y

DRAFT

1. Ethier, Dale

2.

IMOSA PROJECT

FUNCTIONAL REQUIREMENTS

FOR MANAGING INFORMATION IN OFFICE SYSTEM

PREPARED FOR
THE NATIONAL ARCHIVES OF CANADA
IN CONJUNCTION WITH
THE CANADIAN WORKPLACE AUTOMATION RESEARCH CENTRE

BY

DALE ETHIER
DALE R. ETHIER CONSULTING INC.

November 5, 1991

FOREWORD

This report is the culmination of several years of collaborative effort spearheaded by the National Archives of Canada. The other players included Communications Canada, Provenance Systems Inc., and Dale R. Ethier Consulting Inc. The purpose of this part of the IMOSA project was to develop functional requirements for the management of information in office systems across the federal sector. The information management issues and concerns addressed by this report are common to all institutions grappling with the management of corporate information in office systems technologies. From a National Archives perspective, the main goal of these functional requirements is to provide managers with some guidelines on how to manage information in office systems to achieve their respective mandates while adhering to federal policy and legislation governing the management of information in the federal government. Ultimately, the preservation of a corporate memory is of paramount importance to all the players involved for it is the foundation of our knowledge and national identity.

It should be noted that these functional requirements are not standards, but rather guidelines on how technology can serve to address some of the information management problems created by technology itself. Some portions of the requirements have been tested via prototype development, others have not.

The contributions of the IMOSA Functional Requirements Team were indispensable to the production of this report. Their knowledge and experience proved invaluable to the project and we would like to take this opportunity to thank them for their participation.

TEAM MEMBERS

Christiane Desautels	Information Management Standards and Practices
Catherine Zongora	Government Records Branch
Susan Hall	National Archives of Canada
Gilles Legare	Informatics and Records Services
Peter MacKinlay	National Archives of Canada
Paul Marsden	Government Archives Division
Antonio Lechasseur	Historical Resources Branch
	National Archives of Canada
Roy Medaglia	Records Management Services
	Communications Canada

Bruce Miller — Provenance Systems Inc.

Dale Ethier — Dale R. Ethier Consulting Inc.

PROJECT AUTHORITIES

Susan Gillies	Canadian Workplace Automation Research Centre Communications Canada
John McDonald	Director Information Management Standards and Practices Government Records Branch National Archives of Canada

For further information on this report, please contact:

Mr. John McDonald
Director, Information Management
  Standards and Practices Division
Government Records Branch
National Archives of Canada
395 Wellington Street
Ottawa, Ontario
K1A 0N3

TABLE OF CONTENTS

1.0 OVERVIEW

These functional requirements are aimed at those concerned with the management of information in office systems. The rapid developments in technology over the last decade have enabled us to create, receive and manipulate more information than ever. Office systems have placed this power in the hands of the individual. With this power comes a responsibility to ensure that the information of value to the organization in these office systems is properly managed and, where appropriate, preserved for long-term archival and historical reasons. These functional requirements attempt to address these issues.

Within the context of this report, a **USER** refers to those using the system as a management tool in the execution of their daily tasks. The **INFORMATION MANAGER**, or designates, refers to the person charged with the task of carrying out the administrative functions associated with managing the information contained in the system. These tasks equate to the records management tasks of the paper environment. The **APPLICATION MANAGER**, or designates, refers to the person in charge of maintaining the application itself, including such items as assigning field default selections and assigning home prefixes.

2.0 INFORMATION MANAGEMENT CONCEPTS

The information management discipline has traditionally been restricted to the management of information in the form of paper records. However, relatively recent developments in information technology have meant that an increasing proportion of the total records holdings of organizations is being created and stored in electronic form, documents created on personal computers using word processing software, for example. According to current federal government policy, these electronic records must be managed in the same manner, and with the same degree of integrity, as the paper records. However, paper records management practices cannot be directly applied to the management of records in electronic form. The paper handling techniques and practices must therefore be adapted to meet the particular requirements of the electronic environment while adhering to the goals of sound information management practices, that is, the provision of information to users as and when required, while ensuring that obsolete information is disposed of, and that archival information is preserved.

In attempting to adapt paper records management practices to the electronic environment, the National Archives of Canada has examined ways and means of collecting, storing and organizing electronic records in office systems to ensure that they are managed appropriately, along with the corresponding paper records. The functional requirements outlined in this report represent over five years of research into this field. The application described herein will be referred to as the Corporate Information Management Application, or **CIMA**. These innovative functional requirements represent the beginning of a comprehensive solution for the sound management of electronic records.

It should be noted that it is not the mandate of this report to impose recommendations. Rather, it proposes a methodology by which electronic records may be effectively managed. The techniques outlined in this report may or may not eventually find their way into actual working systems. It should also be noted that these functional requirements are based on assumptions made by the contributors, based on their knowledge of the information management discipline.

2.1 INFORMATION MANAGEMENT IN THE FEDERAL SECTOR

In order to understand the functional requirements contained in this report, it is important for the reader to understand certain basic principles regarding the management of paper records and information management in general. The following information is provided for clarification purposes.

SUBJECT CLASSIFICATION SYSTEM PRINCIPLES

The corporate subject classification system forms the core of any Information Management Program. The following basic principles apply to the development and implementation of any such system, whether for paper or electronic information. The system should segregate operational from administrative information to facilitate retention and disposition practices. It should be arranged logically, by main subject groupings, which relate to the program activity structure of the organization. The system should be arranged in a hierarchical manner, progressing from the general to the specific, to facilitate reference and research. The system should be flexible in order to permit additions, modifications and deletions as required. Finally, it should be simple, and easy to use and understand.

The principles of subject classification, applicable to both paper and electronic records, are as follows.

Basic Principle The subject of the information governs the correct selection of the applicable file number.

General to Specific Files should be created on the principle of progression from the general to the more specific. A specific file should not be created before the general file for any one subject, unless it is known that particular subject will grow in the near future. As topics develop, specific related files should be created as required.

Speculative Files

Speculative files, that is, files for which there may be a need in the future but for which no information exists at present, should not be created. The need for a file is governed by the existence of information on that topic.

Mnemonic Aids

Wherever possible, meaningful numbers representing, for example, years, vehicle numbers, etc., should be used as mnemonic aids.

Flexibility

The system should be flexible and allow for the addition and deletion of numbers and subjects. As the quantity of information increases, the need for more subject categories will also increase, and new subjects and files will be required.

Logic

The system should be as simple as possible to facilitate reference and research.

3.0 SYSTEM ARCHITECTURE AND CONFIGURATION

3.1 SYSTEM ARCHITECTURE

The functional requirements of the system described herein are meant for an office system or local area network. All users on a system running such an application should have access to it. Upon calling up the software, the user should be notified of any outstanding BFs (bring forward) or information reservations that were flagged for that particular day (see section 7). They should then be presented with a main menu, some items of which should lead to sub-menus. The particular functions presented to a user will depend on the function assignments granted to the user by the application manager.

An electronic document is identified by its eleven character name under the DOS file-naming conventions. Most applications rely exclusively on DOS to name and locate documents via directories. Within a corporate electronic information management application however, there should be two domains in existence for storage purposes, namely a personal domain and a corporate domain. The personal domain may be defined as a user's personal workspace and may be any or all disk storage space available to a user, including local floppy and internal hard disk storage, and any available space on a shared file server. In this environment, it is usually left up to each individual user to name documents and to decide where to file such documents. The corporate domain is a controlled document storage space which should be administered by the application itself. It is the electronic equivalent to the paper filing system. Documents stored in this environment are under corporate control.

The user's personal domain should be further divided for the purposes of instituting a corporate electronic information management system. A portion of the personal domain should be reserved for the storage of documents submitted to the corporate system. Another portion of a user's personal domain should be devoted to storing other documents. There are no formal processes required for this latter portion. The documents themselves may be categorized into two types. One, documents created by applications such as word processing, spreadsheets, graphics, etc., that are stored in the corporate domain as well as the personal domain. Each of these documents should have associated information such as subject, date filed, document number, version number, that distinguishes them from ordinary electronic documents. Two, documents created by applications such as word processing, spreadsheets, graphics, etc., which are stored in a location of the user's choice, named at the user's discretion. The ability to locate the documents in the second case depends directly on the user's efforts to organize the workspace. No formal filing information is attached or associated with the documents in the second case unless a file number is imbedded in the document itself by the user, again at the user's discretion. The user may move, alter or delete documents at will in this second scenario.

Within the user's personal domain, there should be a distinction made between electronic documents received from others and documents retrieved from a corporate electronic information management system. Documents received usually require some action whereas documents retrieved are usually used for reference purposes or to extract information from the contents. Documents received are usually more urgent whereas documents retrieved are of less concern to the user. The application should therefore differentiate between these two types of documents. This can be accomplished through the creation of two separate directories, that is, Retrieve Directory and Receive Directory.

It is also important for the system to mimic the document dates in the paper environment. The system should automatically record the date a document was filed. It should also accommodate the date of the document itself. In the case of an electronic document, the document date could be the operating system date stamp. In the case of an non-electronic document, this date would have to be entered by the user. The unpredictability of document dates leads to the recommendation to rely only on the date a document was filed. This will eliminate the need for the user to enter all document dates. The system should also record and display for the user the date a document was received in order to assist the user is making decisions as to the appropriate action to be taken concerning a specific document. The system should also be recording the date a document was retrieved from the corporate system in an audit trail in order to trace activity rates on information for the user of the application manager. (see section 17) The average user need not be provided with this information.

3.2 SYSTEM CONFIGURATION

Two types of configuration should exist, namely a personal configuration and a system configuration. The interaction between the corporate system and a user's personal domain is extensive. They rely heavily on each other and are thus interdependent. Each of these two configurations should be conceptualized as two separate entities. The personal configuration should allow the user to modify their home prefix default, login password, and their receive and retrieve directories. All of these functions should also be available to the application manager as well as to the user.

The system configuration should be comprised of sub-functions. The application manager should be able to define fields as mandatory or non-mandatory and default field contents as per the needs of the organization. When configuring the system, the application manager should be presented with two options per field, yes or no, with the default being non-mandatory. The system should also allow the selection of voluntary or involuntary collection of information. The default for this option should be set at voluntary.

The application manager should also regulate the type of documents that may be included in the system. The opportunity to create a scrollable list of document types should be made available to the application manager. From this option, the application manager should be able to add a new document type, edit an existing document type, or delete an existing document type. When adding or editing a document type, the application manager should be presented with a screen to enter the appropriate information including the name of the document type. If editing, the system should display the existing data on that particular type. All fields presented should be editable by the application manager. If entering a new document type, the application manager should be presented with a blank screen to enter the document type name. Once the new document type is accepted, the system should assign a unique document type identifier to the new entry and add it to the other existing document type information. All changes should be prefaced by a confirmation prompt.

3.3 USER GROUP MANAGEMENT

This section deals with the operations involving the creation and management of system users and groups of system users. The group scenario should exist to facilitate the assignment of functions to more than one user at a time. Each user or group should be given access to one or more prefixes. Within each prefix, access should be limited to a range of subject numbers, the default being all numbers contained within that particular prefix.

Each user or group should be assigned functions to which they have access. The default here should be very restrictive. Each user or group should have a minimum of one function, namely the ability to login to the application. This should be made the default for the assignment of access to the system functions.

The system must allow for the creation of users. This function, restricted to the application manager, should include the following fields: user name, notes, login password for the user, user's security clearance level, home prefix, other prefixes to which the user has access, file range within each prefix to which the user has access, language, receive directory, retrieve directory, and accessible functions. The application should also allow the editing of user information and the deletion of a user and all associated information. The functions involved in the creation of a group are essentially the same as those for the creation of a user, except that a group name should be specified, along with a list of its members.

3.4 FILING SYSTEM ORGANIZATION

Due to the large number of different approaches to information management throughout the federal sector, no single method of filing system organization is presently being used to manage information across government. It is estimated that approximately 85% of federal institutions are currently using the block numeric subject classification system to index their paper-based information resources. These functional requirements have therefore been geared to accommodate the rules inherent in the block numeric system. Within this structure, each institution has adapted the principles to meet their own requirements. Therefore, although the subject classification systems have a common base, many permutations are in existence. Numerous variations are currently in use which have evolved over time to meet each institution's particular needs. The filing system organization inherent in the software should therefore be flexible enough to accommodate a number of scenarios. Research into this area has shown that five main permutations of the block numeric system are possible. These are outlined below.

Parallel Systems

An organization running parallel block numeric subject classification systems uses the prefix to represent totally different filing systems unique to particular organizational units. Each subject and file title under one prefix is independent of those under another prefix. It is estimated that approximately 5% of institutions in the federal sector are operating with this scenario.

Program-Based Systems

A program-based use of the block numeric system involves the use of a prefix to represent a unique program activity. All subjects and file titles are unique to each prefix, the premise being that all related topics are clustered under a common prefix representing the formal activity to which they relate. Numbers representing subjects and files may be used freely within each prefix. It is estimated that approximately 20% of institutions in the federal sector are operating with this scenario.

Location-Coded Systems

In this scheme, the prefix is used to denote the location of information. Subjects are uniform throughout all prefixes. File titles may differ within each prefix, that is, at the file level, each prefix's titles are completely independent. Numbers representing subjects must be uniform across prefixes but numbers representing files may be used freely across different prefixes. It is estimated that approximately 50% of institutions in the federal sector are operating high level menu which allows the user to select what function they want to use. All application administrative functions should be handled by the login function. Once logged into the main office system, the user should be asked to log into the electronic information management application by specifying a user name and password. The system should allow only three repeat attempts on an incorrect password, after which the system should reject any further attempts under that particular user name. The user should only be shown the functions to which access has been granted by the application manager. The login function should either spawn or execute functions depending on memory requirements. The system should verify that sufficient memory is available prior to proceeding.

The system should then automatically execute the following functions. The system should check for BFs (bring forward) for that user name. If such BF items exist, the system should flag the user with a message. The system should also check to determine if any previously filed documents were re-filed and notify the user via message providing the user with the document and version numbers, and the old and new file numbers.

The personal information on the user such as preferred language, personal security level and home prefix, should be passed to the invoked functions automatically and should be transparent to the user. Some system information should also be passed, monitor type for example. Other invisible tasks should also be executed such as clear screen and draw screen.

3.7 HELP FUNCTION

The help function should be context sensitive and should be displayed in the language preference setting chosen by the user. A single key should be dedicated to the help function throughout the application. Regardless of the location of the user in the menu hierarchy, depressing the help key should result in the display of a help message which describes the screen, fields and menu items.

When the help key is depressed, the system should reference an index to a help file. When the associated message is located, the application should display it on the screen. If no message exists for a particular field or menu item, then the system should default to the help message for the screen. Help messages should exist for all fields and menu items and the system should never display a blank screen as an acceptable result of a user request for help.

Separate english and french messages should exist for each of the top level functions associated with the system. The help files should be stored in a separate directory along with all of the other application files. The with this scenario.

Organization-Based Systems

An organization-based use of the block numeric system is adopted when an organization is divided into distinct groups of users. A prefix is then assigned for each distinct group, usually based on organization structure. A single, organization-wide set of numbers and corresponding subjects and files is created as a "master list" from which titles are selected on an as required basis for each organization and the corresponding prefix is applied to denote location and ownership. It is estimated that approximately 20% of institutions in the federal sector are operating with this scenario.

Systems Without a Prefix

In some instances, especially in small, transitory organizations, there is no need to use prefixes at all. In this scenario, an artificial prefix would be applied, the organization's acronym for example, and all users would be provided access to the prefix and would treat it as if it were invisible. It is estimated that approximately 5% of institutions in the federal sector are operating with this scenario.

Several key points must be taken into account if the system is to accommodate the five permutations of the block numeric subject classification system as outlined above. It is essential that, in each case, the prefix be treated as an inseparable part of the file number. In three of the five scenarios above, the file number without the prefix is totally meaningless. It is also necessary that, with few exceptions, all users within an institution may view all file **titles** within that organization. The system should not restrict file title access, but should restrict access to the **contents** of files. The system should display a complete list of prefixes and their descriptions. Access to prefixes to which a user does not have access should be denied.

In some of the five variations on the block numeric subject classification system, considerable overlap exists between prefixes. In order to avoid duplication of effort during data entry and/or classification system re-design, the software should allow for the replication of sections, subjects and files from one prefix to another, or to multiple prefixes.

3.6 LOGIN FUNCTION

The login function should be invoked when a user wants access to the application. Login should also be responsible for the management of the extensions of all help file names should be unique, "hlp" for example, to promote easy identification. The text of the help messages should be in paragraph form with a wrap around feature, and should be scrollable where more than one screen of information is necessary. The existence of additional information, beyond a single screen, should be indicated and obvious to the user. Movement in the help screens should be accomplished via the cursor movement keys.

3.8 USER INTERFACE

The user interface of the application will no doubt be different from application to application but a few guidelines would perhaps be useful. Based on past experience with these kinds of systems, some information on user interface is provided below.

The user interface should be simple and easy to use. The user should input with the fewest possible keystrokes. Novice users should be able to quickly determine which keys execute which functions. More advanced users should be able to quickly and directly execute desired functions without being slowed by overly cumbersome instructions or steps. Expert users should be provided with an expert mode that would allow them to bypass much of the interface. Lastly, the user interface characteristics should be applied uniformly throughout the application.

3.9 GENERAL SYSTEM REQUIREMENTS

The following general system requirements are recommended in the design of a user friendly application for use by a wide range of individuals with varying degrees of automation skills.

1. Menu items should be ordered by frequency of use.

2. Prefix should always default to user's home prefix. The prefix and file number are inseparable and should always appear together.

3. Each field should be designated as either search or scan type fields. Search fields are typically used for numerical or text fields and should result in immediate response. Scan fields are used for text and query time can be unpredictably long and slow.
4. When information is needed on a particular field, a single function key should be dedicated to bringing up a help menu explaining the purpose and parameters of the given field.

5. It should be apparent to the user which fields are active and which are inactive, and of the active fields, which are mandatory and which are optional.

6. Standard date and time formats should be established based on user familiarity and needs.

7. The application should be case insensitive.

4.0 AUTHENTICITY AND VERIFICATION ISSUES

It is important for users to have on-screen verification that a document has been "signed" by the author, and that it has not been altered since the time of filing. In the paper environment, the signature is the unique mark of authenticity for a document. An equally unique method of validation is required in an electronic environment. The greater the technical integrity of such a method, the more confidence users will place in it. Since an author's signature is impractical in an electronic environment at this time, other approaches, such as a magnetic stripe reader, are needed if the requirement is serious enough to warrant the expenditure.

The application software should incorporate authentication into its program. For example, each document profile could contain two special fields, namely AUTHENTICATED and AUTHENTICATED DATE. The first field would simply be a yes or no option and would automatically be marked yes if, on filing the author's document was authenticated by equipment such as a magnetic stripe reader. Since users retrieve only copies of documents, not originals, the retrieval of a copy to a user's personal workspace would have no impact on authenticity. If alterations are made to the copy, the document would become a new version or, if the subject has changed, a new document. The new document or version would undergo authentication during re-filing. The software must ensure that no authenticated document is ever altered under any circumstances. Permission to alter documents should be denied to anyone, including the application manager and the information manager.

4.1 THE VERIFICATION PROCESS

The verification function is carried out by the information manager. As often happens in the paper environment, the file numbers assigned to documents by the average user are wrong up to 40% of the time. In an electronic document management application, this could spell disaster. Therefore, the information manager must be able to verify the file numbers assigned to documents by users. The verification process itself involves noting the author and/or recipient of the document, scanning the document to determine the subject, identifying the relevant section in the filing system, narrowing down the options to a specific file number, verifying that the number is correct, or if it is incorrect, selecting and entering the appropriate file number. In the paper environment, approximately forty documents can be verified per hour, per operator. It is imperative that the system be able to keep up with or surpass this rate.

In order to execute this function, the information manager should be presented with a scrollable list of documents to be verified in the order in which they were submitted. A document should then be selected from the list for verification. The following fields should be included in order to carry out the verification.

Prefix - editable
File number - editable, the look up function should be available for this purpose
To and From - display only, should not be able to edit these fields
Subject - display only, should not be able to edit this field
Cross-reference - editable, this field is never entered by the user, should only completed at verify time by the information manager
Summary - editable
Essential - non-mandatory, option to choose yes or no
Security level - defaulted to unclassified, option to select protected, confidential, secret and top secret should be provided
Document number - display only, not editable
Version number - display only, not editable

If a file number is changed, the system should confirm the change and move the document to its new destination. Any BFs or charge outs should be updated to reflect the change. Once the document has been verified, it should be noted in the document profile. A field for verification with a yes/no option, automatically completed by the system would suffice.

5.0 FORMAT STANDARDIZATION

5.1 UNIVERSAL DOCUMENT FORMAT

In order to facilitate the management of information in a CIMA, it is recommended that electronic documents be converted from native format to a universal document format. For example, the original application software in which a document was created may no longer exist, and the difficulty of converting between native formats may present problems to the application manager. All documents should therefore be converted to a universal format as soon as possible after initial creation and filing. The ideal time to accomplish this is at the time of filing. In this way, no native formats would ever stored in the filing system, and all documents would be in a single, uniform format. Users would be able to convert to a native format of choice upon retrieval of a document from the filing system in order to work with that document.

The conversion to a universal document format is considered essential for managing electronic documents in a modern office environment for the following reasons:

1) It is likely that in a large organization, many different native document formats will be in existence. Each unique software application used to create electronic documents equates to a different native format. The greater the number of unique applications, the greater the difficulty in managing the documents.

2) All native format generally changes at least once per year with new software releases. Some may also become obsolete with time, and therefore become unreadable. It is technologically difficult to keep pace with manufacturers' software.application changes.

3) If documents are maintained in native format, the CIMA users may view them but would have to have access to the native applications used to create them and be trained in the operation of that native application in order to work with the document.

4) Converting multiple native formats to other such formats to avoid obsolescence would be expensive and time consuming.

Clearly, it is essential that, if technologically possible, all electronic documents should be maintained in the CIMA in a single, universal format. The advantages of this strategy would include the following points.

1) The originating agency would only have to maintain a single document format, thus reducing cost of maintaining multiple formats in terms of dollars and human resources.

2) Users would not be able to alter documents stored in the CIMA. Rather, at retrieval time, a user would request that a document be supplied in a native format of choice. The document would then be converted from the universal format to the specified native format for editing purposes. This would help ensure the integrity of the information stored in the corporate system.

3) The host organization could offer high-quality, full-function document viewing and editing features for the universal document format.

4) Reviewing large quantities of electronic documents could be accomplished by a single process. The use of a universal format would also assist the National Archives with the assessment of large quantities of documents via the use of a single process instead of different, often outdated processes for each unique proprietary document storage format.

One of the universal formats now being considered throughout North America, ODA/ODIF (Office Document Architecture/Office Document Interchange Format), holds promise for the CIMA. It is the format that is currently being considered by the Treasury Board Secretariat and is now becoming commercially viable. ODA/ODIF has a reputation for being the most comprehensive of all universal formats to date. It accounts for the inclusion of non-textual (graphics, voice, etc.) components of documents as well as textual and is a true compound document architecture which is important for documents stored in image form.

According to a report by Strategic Technologies Inc. entitled Recommendations for a Compound Document Representation Format for the Canadian Patent Office, dated February 1989, the alternative to ODA/ODIF is SGML (Standard Graphics Markup Language). This format is popular in the United States but has some disadvantages. It is not a true compound document architecture. It is designed as a mark-up language where ODA/ODIF was designed for document interchange. Also, SGML has no mechanism for encoding pictorial elements, an important deficiency.

It should be technically feasible to incorporate a universal document format into CIMA technology in the near future.

6.0 PERSONAL DOCUMENT MANAGEMENT

The personal document management function is the vehicle by which a user would re-submit a document to the system. It should be the only point which will accommodate the re-submission of documents. The nature of the management of electronic documents dictates that specific, controlled conditions be set up within personal workspaces. Certain requirements exist for the introduction of such software. A set of assumptions has been developed in order to create the personal document management requirements and is summarized in this chapter.

6.1 PREMISES

It is presumed that users differentiate between documents received from others and documents retrieved from an electronic information management system. Documents received usually require some action whereas documents retrieved are usually used for reference purposes or to extract information from the contents. Documents received are usually more urgent where documents retrieved are of less concern to the user. The application should therefore differentiate between these two types of documents. This can be accomplished through the creation of two separate directories, that is, Retrieve Directory and Receive Directory. It is estimated that a typical user would have between ten and twenty electronic documents in their receive and retrieve directories at any one time. It is assumed that only documents contained in these dedicated directories would be recognized by the system. Electronic documents stored outside these directories would not be retrievable via the system. Also, the documents in the dedicated directories must maintain their profile information. Without the profiles, the documents would not be recognized by the software.

It is presumed that users would most often invoke the personal document management function to work on a document. It is estimated that the average user would work on a given document three or four times prior to re-submitting the document. Therefore, the system should allow documents that are being worked on by a user to be kept over an indefinite period of time, even if the user exits the personal document management feature. It is also recognized that users rely heavily on the document's subject line or profile to determine the action to be taken regarding a document. The subject line stimulates the user's memory and helps to determine what action will be taken on the document. Therefore, it is insufficient for the user to be provided with only the DOS filename for each document in their personal domain. The following information is required to help the user assess the importance and action required for each document. The subject, document number, version number, and the date filed should all become part of the document profile. When the profiles are displayed, they should be in document number order, and version number within.

In addition, when users are viewing the list of documents in their personal workspaces, they will want to be able to work on a document without leaving the application. A method of invoking the application in which that document was created will be necessary. The software should be capable of automatically launching the proper application. Users also need to know whether or not a document they are working on has been previously filed in the automated system. A method of flagging these documents is therefore required.

It is presumed that average users do not often concern themselves with the rules of DOS path and directory management. Most users are content to work out of a few directories. Therefore, moving documents between directories should not be a major concern. Moving documents should therefore be accomplished by using DOS directly rather than creating this capability within the application. Also, there is a requirement for users to be able to delete documents. This option should also be integrated into the system.

All documents should have unique operating system names that meet the following criteria: each document must be uniquely numbered, and must have a version number. The operating system file name should never be changed. If it is changed, the system would not be able to recognize that document. The operating system name would act as the link between the document and the system. The parameters established for the total number of documents and subsequent versions must be large. A suggested maximum of 100,000,000 documents with a possibility of 99 versions for each is one option.

Users would also require the ability to execute multi-session editing of documents. In other words, a user should be able to edit a document over an indefinite period of time, modifying as required and as often as required, without restriction.

6.2 THE FILING PROCESS

Users should be presented with a choice of working on documents received, documents retrieved or both, which would yield all documents. Once selected, the system should display a list including the document profiles for each one. The document profile should include the subject of the document, document number, version number, and the date the document was filed. Ideally, the document subject should appear in its entirety but constraints may require truncation of this field. The sorting order of this list should be by document retrieved or received, then by document number and by version number within. For each document listed, whether or not the document has already been worked on by the user should be indicated by a flag.

Three options should then be made available to the user. The option to delete the document, followed by a confirmation message should be available. Second, the ability to work on a document should be included. Third, the ability to view a detailed profile of each document is desirable. The detailed profile should include the following mandatory, editable fields: to, from, security level, document type, subject, and summary. These fields must be editable if the document profile is to be kept current. This is especially important if the user has edited the document and wants to re-submit it. Consideration should also be given to providing fixed lists for these fields. In addition, fields for prefix, file number, date the document was filed, who it was filed by, the document number and version number, and the storage medium should also be included in the detailed profile. Long text fields such as subject and summary should be horizontally scrollable or expandable. The date filed, filed by, and version number fields should automatically be updated by the system. All of the characteristics, including mandatory, editable, fixed list availability, scrollable or expandable, and automatic update, which are specific to certain fields should be made apparent to the user. The user should be able to move from one document to the next without returning to the list of document profiles.

6.3 SUB-FUNCTIONS

Once a document has been selected from the list by the user, the following functions should then be made available.

Work On Function

This function should permit the user to work on the chosen document in the original application in which the document was created. It should restrict alteration of the operating system name of the document since this name is its only link to the system. The operating system name should be protected or a warning message should be issued to the user if alteration is attempted.

Quick View Function

This function should allow the user to view the document selected without having to load the original application.

Submit Function

The submit function should permit the user to file a document using information provided in the profile of the original document. The user should be able to retain a copy of the original document if desired. If a copy is kept, it should be stored in the dedicated retrieve directory. Also, a user should be able to file <u>into</u> any file, regardless of security level.

Quick Print Function

The user should be able to print a rendition of the selected document without having to load the original application.

Delete and Exit Functions

Users should be permitted to delete a document from their dedicated directories and they should be prompted for confirmation of this action prior to execution. The user should of course be able to exit the detailed profile or return to the beginning of the selection process.

7.0 APPLICATION MANAGEMENT

7.1 DOCUMENT MAINTENANCE

Once a document has been submitted, the information manager may need to alter the document profile, even after verification. As a document is submitted to the system, it should pass through three stages, namely submission, verification and post-verification. During submission, the user would enter essential information and perhaps some optional data. During verification, the information manager would verify the classification of the document, move it if necessary, and enter additional fields of minimal interest to the average user. Typical changes would include entering cross-reference information, or updating the subject line. In the post-verification stage, when the action on the document is supposedly complete, normally there should be no reason to alter it. However, circumstances change and the information manager may need to change some of the information associated with the document. For example, the security level may need alteration as a result of a security policy change, a document may become essential, or additional cross-references may need to be added. The system should therefore allow modifications to the document profile after submission and verification are complete. However, certain fields must be protected from alteration at all times including FROM, TO, DATE FILED, and FILED BY.

The document maintenance function should therefore have two sub-functions; edit the document profile, and move a document. In editing a document profile, the protected fields should be apparent to the information manager. The system should ask the information manager to confirm all changes to the profile prior to acceptance. The information manager should be able to invoke the look up function from Document Maintenance in order to verify or look up file numbers. If a file number is changed, the system should adjust the version number to one and notify the author of the change. The system should also allow viewing of the document profile in full and the printing of the profile and the document, if desired.

7.2 FILE MAINTENANCE

The File Maintenance option should allow the information manager to create a new file and complete its profile, delete a file and its contents, modify an existing file's profile, delete the contents of a file, move the contents of a file, and look up file numbers. All of these functions apply to both electronic documents and paper files with the exception of delete the contents of a file, delete a file and its contents, and move the contents of a file, where only the documents in the electronic file are manipulated. The equivalent actions should manually be carried out in the corresponding paper files. These changes are not included in the application.

Create Function

When creating a new file, the information manager should be presented with the file profile fields outlined below.

Prefix
File number - cannot be modified after file creation - the move function should be used for this purpose
Title (english), Titre (french)
Security level - defaulted to unclassified, option for protected, confidential, secret and top secret should be provided as a minimum
Comments - free text field for any comments relating to the file
Cross-reference - for related file numbers and explanatory text
Creation date - cannot be edited
Status - open or closed
Personal Information Bank number - eight characters, non-mandatory
Archival value - defaulted to no, option for yes
Active/dormant - number of years file is to remain active, number of years file is to be dormant prior to disposition
Disposition date - date that disposition action should take place
Disposition action - the manner of final disposition, either keep or destroy
Review date - date the file is to be reviewed on
Subject description - english, at the subject level only
Description de sujet - french, at the subject level only

When assigning a file number, the application should allow a wide range of permutations. For example, a file number will always be composed of a prefix; a subject number, usually four digits; at least one, and up to four sub-file numbers. The prefix should be able to accommodate from one to six characters, including numbers, alphabetical characters and symbols. The subject and sub-file levels should be able to accommodate up to twelve characters each. Given these parameters, the file input field could be defaulted to a maximum of approximately fifty characters since the likelihood of a file number reaching its maximum allowable length is highly unlikely. The description fields are associated with the subject level in the filing system hierarchy. This field should be presented to the user for information purposes but should not be editable except by the Information Manager for filing system changes.

After completion of these fields, all of which should be mandatory with the exception of comments, cross-reference, personal information bank number, disposition date, review date, and description, the system should automatically create the first volume profile and fill out the necessary information including the volume number, the date the volume was created, that is the present date, the date the of the first piece of correspondence on the file (defaulted to the present date), the date of the last piece of correspondence (left blank until the volume is closed), the status of the file (defaulted to open), the security level of the volume (defaulted to unclassified) and the location of the volume (defaulted to the prefix location). The creation of additional volumes of the same file should be accomplished through the Records Centre Management function (see below).

The information manager should then be given the option to accept or reject the new file. The system should perform a validation to determine if the file number is a valid one and confirm acceptance of the newly created file.

The information manager should also be able to delete or remove a file and all of the documents it contains. In order for the system to delete a file, the file profile, all document profiles within that file and the documents themselves would have to be deleted. A rule that no file can be destroyed unless all of its sub-files have previously been destroyed should also be in place. The system should request the full file number to be deleted and confirm the action prior to proceeding. The system should also verify that no sub-files exist prior to the deletion. If sub-files do exist, a message to that effect should be presented on the screen and the action aborted. If a file that is being deleted has an outstanding BF associated with it, the system should notify the information manager who would have the option to proceed or cancel the action. If the information manager decides to proceed, the system should cancel the BF and notify the user who requested the BF of the file deletion.

In order to accommodate the many changes that may occur in the filing of documents, the system should allow the information manager to move the contents of a file to another file. There should be the option to move a group of selected documents, all documents, a range of documents or a single document. The system should prompt the user for the source file number and the destination file number and should validate that both exist. If the source file does not exist, the system should display and error message. However, if the destination file does not exist, the system should ask if it should be created. If yes, the file create scenario should be repeated with the exception that the file number should be inserted by the system and not be editable. Once the move is accepted, the system should confirm the action prior to execution.

7.3 FILING SYSTEM MAINTENANCE

Filing System Maintenance corresponds to the some of the more traditional functions associated with the maintenance of a subject classification system in the paper world. The system must allow the information manager to create, delete or edit a prefix or section of the filing system. Although this is a necessary and important function, amendments to this level of the filing system should occur infrequently with the exception of a major filing system modification.

Because the changes involved in filing system maintenance directly affect the structure of the filing system itself, certain rules should be established to provide a controlled framework for change. The hierarchical structure of the filing system should be based on the subject classification system of the paper environment within the organization, in this case, the block numeric subject classification system (see section 2). In a sense, one should mirror the other. As such, it is imperative that the rules governing filing system maintenance be adopted and adhered to in the electronic system.

Prefix Management

Each prefix in the system must be unique, and must correspond to a unique prefix name. The system must allow the deletion and modification of prefixes by the assigned information manager within the organization. This function should not be made available to the average user. Since the prefix represents the top level within the filing system hierarchy, a change to the prefix level must be reflected at all corresponding lower levels in the hierarchy. The system should therefore allow global modification at the prefix level. It should also be possible for the Information Manager to move sections of a given prefix to another prefix. When such a change occurs, it is imperative that the system ensure that all information related to that section is moved as well.

The deletion of a prefix should also be possible although this action would likely occur infrequently. The system should allow the deletion of a prefix only if it is confirmed that there are no references to that prefix in any of the data remaining in the system. If data still exist, the data would have to be moved to another prefix prior to deletion.

In order to access the prefix management function in the system, the Information Manager should be presented with a scrollable list of available prefixes, sorted in alphabetical or numerical order. From this scrollable list, the information manager should be able to execute the functions of adding, deleting or editing any of the prefixes. Upon completion of any of these actions, the system should confirm the acceptance of the actions prior to executing the changes as a safeguard against errors. One option is to force the Information Manager to confirm the changes by making acceptance of the modifications a separate, additional function. The system should also verify that the proposed changes are valid; that the changes do not conflict with existing prefix information contained in the system. Because the changes at this high level in the filing system hierarchy could mean massive data changes, execution of this function could be extremely time consuming. A default of a specified time period should be established, perhaps one or two minutes depending on the size of the database, at which time the information manager would be given the option to continue or abort the changes.

Section Management

Within each prefix, sections representing major subject areas exist. Within each section, specific subjects exist. Each section must span a unique range of numbers within a particular prefix, and each subject within a section must be associated with a unique identifier. These rules are very similar to those of the paper system (see section 2). The standard used in the block numeric subject classification system is a four digit number, from 0001 to 9999, allowing a full 10,000 number range to be used for the identification of different subjects within the organization. This may vary from a two or three digit number to a five digit number, depending on the size of the organization and/or the number of different subjects that must be accounted for in the filing system.

Regardless of the length of number being used, each section should be named and assigned a unique range of numbers under each prefix. The upper limit of each range of numbers should be recorded by the system, implying that those subjects after the last section, up to the upper limit set for that section, are members of that section. The table below illustrates this simple but somewhat confusing concept.

Section	Upper Limit	Subject Number Range
1	1000	0001 to 1000
2	2000	1001 to 2000
3	5000	2001 to 5000

It should be noted that the numbers should be continuous, without gaps, but should not be restricted as to length. It should also be noted that the lower bound should be fixed at 0001, but that the upper bound should float, depending on organizational needs. Flexibility should be built into the system to allow an upper limit of 12 digits, even though the most common usage will likely be limited to four digits.

The system should allow any new section to be inserted as long as the new section's upper limit is greater than the previous section's upper limit. It should also be possible to delete any section and the remaining sections should automatically adjust to reflect the deletion. The system should not allow the lowest possible number, that is 0001, to be entered as a section's upper limit. If an upper limit is modified, the system should not accept a new number unless it is greater than the previous section's upper limit. In parallel to this concept, the modified upper limit should be less than the next section's upper limit. The system should not allow any deviation from this rule.

In order to carry out any of the above-described functions at the section level, the Information Manager should be presented with a scrollable list of sections within a pre-selected prefix. The list should appear sorted in numerical order by section upper limit. From this list, the Information Manager should be able to edit, delete or create a section. The name of the section, its upper limit, and a brief description of the section should be mandatory for each section in the system. Whether editing, deleting or creating a section, the system should prompt for acceptance of the data and should provide the user with the option to abandon the changes made and return to the section list.

7.4 BF (BRING FORWARD) MAINTENANCE

This function would allow the user to request that either an electronic document, a paper file, or both, be brought forward at a specified future date. The BF function should be used exclusively by the user, not the information manager. The only involvement by the information manager would be when processing the BFs for paper files (see 7.5, Records Centre Management below).

The two most common reasons for a BF request are for follow-up or for action. Both involve recalling a document at a specified future date. The BF of an electronic document is somewhat different from the traditional BF of a paper file. Rather than providing the user with the actual BFd documents, the system should simply notify the user who would locate and retrieve the documents at their convenience. Copying entire electronic documents to a user's personal workspace would not be practical due to the potential number of documents that could be involved. The amount of disk space required could be quite substantial.

There is another difference between the BF of an electronic document and a paper file. A BFd paper file is charged out to the requesting user and is therefore unavailable for other users until it is returned. The BF of an electronic document would only involve user notification. When the user wanted to view or work on the BFd document, the user would retrieve that document, in fact copying it into their personal workspace, leaving the original in place for the use of others.

A user should be able to request the BF of a document within or outside the user's home prefix. Because most BF entries will be made within a user's home prefix, the system should be defaulted to the home prefix. However, the user should be able to override this information and enter any prefix to which access has been granted by the application manager. A user should also be able to request the BF of a paper file alone or to accompany an electronic document through the application. In either case, a directive to the information manager should be issued instructing the forwarding of the file at the appropriate time. The information manager would then proceed with the processing of the BF on the specified date (see 7.5, Records Centre Management below). Whenever a user specifies that a document or file should be brought forward at a specified future date, the system should create a BF entry. The BF entry would become the record of the request with all associated details.

Several underlying premises affect the BF function in such an application. There is an average of approximately five BFs per user per year but in certain instances, because of unique work functions, a user must be able to enter an unlimited number of BFs. Also, there should be no default period for a BF. It is impossible to predict an average time period for this function. Therefore, the system should prompt the user for the BF date every time the user enters a request. The due date for a BF is usually the BF date minus one day. In other words,, the user notification in the case of an electronic document, or the information manager notification in the case of a paper file, should be issued one day before the actual BF date. The BF date itself should be expressed as a calendar date rather than an elapsed period of time in order to conform with all date fields contained in the system.

In addition to these points, there is likely to be a limited requirement for the user to be able to specify the version number of an electronic document when entering a BF. The user should know that between the time that the BF is entered and the delivery of the actual BF reminder, the version may have been incremented several times by other users. Therefore, a user should not assume that by entering a version number, the latest version will be the one that the system BFs at reminder time. In order to avoid misleading the user, the system should not default to the latest version number. Rather the version number here should not be an issue.

The user will also be likely to BF documents bearing the present date most often. However, the system should allow the user to BF a document with a date sometime in the past. Therefore, the user must be able to specify the target date (see Target Date below). In the case of a user requesting an electronic document and its accompanying paper file, the application should be able to complete the target date by using the electronic document's date. In the case of a user requesting a paper file only, the system should prompt the user for the desired target date.

Target Date

In order to process a BF request for a paper file, there must be a means of determining the correct volume to deliver to the user. In the paper environment, a user specifies a date or date range. If the date is not specified, it is interpreted to mean that the user is looking for the current volume of a file. The application should introduce the concept of a "target date" as the mechanism by which a user could specify the date range of the desired information. Whenever a user requests a BF, the approximate or exact date of the requested information should be specified. The proper volume would then be identified based on the target date.

At login time, a user should automatically be notified of any BFs that came due that day and any BFs from previous days that have yet to be processed. The user should also be able to review BF notifications at any time during the day. The system should allow this by permitting the user to view a sequential list of outstanding BF notifications. The document number, version number, subject and BF date should be provided for each. Also, the option to delete the notification should be made available.

When a user enters a BF request, the system should offer three options, namely to BF an electronic document with an option to include the paper file, BF a paper file only, or review current BF notifications. By choosing either of the first two selections, the user should be presented with a screen containing fields for the document number, the version number, desired BF date, whether or not the corresponding paper file is also requested, and the target date. The document number, and BF date fields should be mandatory. The version number field, BF date field and target date fields should be blank and editable, and the option to request the paper file field should be defaulted to 'yes' with the option to change to 'no'. If the user does not want the paper file included in the BF, the target date would be ignored and left blank. When requesting the BF of a paper file only, the user should be required to enter a file number, a target date, and a BF date only. The user should also be able to delete a BF request at any time.

Once the information has been entered the user should be given the option to accept the request as defined, thus creating a valid BF entry, or clear the fields to begin again, or exit the BF entry function.

7.5 RECORDS CENTRE MANAGEMENT

The functions associated with Records Centre Management are designed to assist in the management of traditional paper records. These functions may be considered as a series of file issue control functions which are part of the functions of a traditional automated records management system. A series of four interrelated functions exist in this area, namely the charge-out, bring-forward, volume maintenance, and report generation functions. The information manager will be the primary user of this part of the system.

The charge-out function involves keeping information on the data required to process the charging out of a paper file. The system should allow the information manager to add, delete and edit file charge-out information at will. The information manager should also be able to print file recall notices via this function. For each file that is to be charged out, the system should record the name of the user to whom the file is being charged, prefix, file number, volume number, date the file was charged out, and recall date. A scrollable list of user names should be presented to the information manager in alphabetical order for selection purposes. Depending on the size of the organization, adequate navigation tools should be implemented to allow the information manager to locate a name efficiently. The information manager should also be able to enter a name which does not appear on the user list. The prefix field should default to the user's home prefix, once selected. A scrollable list of valid prefixes should also be available from which other selections could be made if desired. The file number field should be mandatory and the look up function should be invoked from this position (see section 11). A fixed list selection should also be made available for the selection of volume number. A valid list of volume numbers for the selected file number should appear. The date the file is being charged out should also be mandatory and editable, and the default should be set at the present date. The file recall date is the date that the file is to be returned. This field should automatically be filled in by the system and the default should be based on organizational need. Generally speaking, a thirty day period of file use is satisfactory in most cases. The recall date default should therefore be thirty days following the date the file was charged out to the user. A recall date should be associated with every charge-out executed and should also be editable by the information manager in case a file should have to be recalled earlier or later than the defaulted thirty day limit. The information manager should be given the option to instruct the system to accept the charge-out as entered, delete a charge-out completely with a confirmation prompt, print a recall notice for the selected charge-out entry, clear all fields so that new data can be entered, and exit the charge-out function completely.

BF (Bring Forward) Function

This section applies only to those BF entries that request a paper file. The BF function should be totally user driven. The user would enter and maintain all BF requests. The information manager would process only those BFs that call for paper files, either paper files only or paper files accompanying an electronic document. The processing to be done by the information manager would involve charging out the file on the appropriate date. The information manager should also be able to cancel a BF if it is known that the user wishes to cancel a BF request. There is no requirement for the information manager to edit a BF request since the process is user driven. The information manager would, however, want to be able to print a report in the form of a comprehensive listing of outstanding BFs of paper files by date in order to physically retrieve those files.

Two assumptions, based on discussions with several users, have been made in the context of this section of the functional requirements. When a user wants to use the BF function, it is either for a specific document, or for the entire file. Often when a user requests an electronic document, the user would also want to see the corresponding paper file for reference purposes. These two assumptions are at the heart of the BF functional requirement definition and should be kept in mind throughout this section.

In order to process a BF for a paper file, the information manager needs to know several pieces of information about the file including the date of the BF, the file number, the volume number, the user's name, and the file title.

The Records Centre Management portion of the BF function should be straight forward. Because it is user driven, the only action involved should be the actual processing of the requests for paper files. This portion of the system should therefore be list oriented. The list of BFs could contain none, one or multiple entries at any given time. Because of this list management orientation, the information manager should be presented with a list of BF entries available off the main records centre management menu. The list should include BF date, prefix, file and volume numbers and name of the user requesting the file. The list should be sorted by BF date, then by prefix, file and volume number. The list should only contain BFs that meet a specified date, usually the present date or the next day's date. This should be predetermined by the needs and practices of the organization. The information manager should be able to select an entry from the list for viewing or editing. When a particular entry is selected, the information manager should be presented with all pertinent information on that particular BF request. The information manager should also be able to delete a BF request for a paper file if it is known that the user wants to cancel the BF.

In addition to pertinent information on a selected BF entry, the information manager should be presented with a date field defaulted to the present date, which would allow the information manager to charge out the file. A file recall date, based on an established time period, should automatically be assigned by the system, usually thirty days from the present date. The information manager should also be able to override this date and enter any valid future date in this field. Once a BF has been processed or charged out, the system should delete that entry from the BF list so that it will not appear on the BF entry list again. Once the information manager has completed the charge out action, the system should transfer that data to a charge out entry.

Volume Maintenance

This section refers to the start and end dates of physical volumes of paper files. It is important to the BF function in that the user must be able to pinpoint a date or target a date range of information when searching for information or entering a BF date. Since the user may want the corresponding paper file, the system should be able to track date ranges of volumes of paper files. At the initial paper file creation time, each file begins at volume one. As the amount of material increases, additional sequential volumes are created by the information manager as the physical need dictates.

The system should recognize the sequential nature of volumes and should also recognize that there is not always a complete set of volumes for a given file. For example, some older volumes could have been destroyed under a disposition authority, so the first existing volume number for a particular file may not be volume one. When a volume is created, a label containing file number, volume number, and subject title should be produced for application to the paper file jacket. Each organization is likely to have different label requirements depending on preference, and on file jackets. The system should be flexible enough to allow label format changes to accommodate changing needs.

A specific set of rules applies to volume maintenance in the paper environment that should by considered when developing or implementing a CIMA. These rules are outlined below.

RULES

1. Volumes are always sequential but do not always begin at volume one due to scheduling activities.

2. All volumes but the current volume must be closed.

3. When a volume is created, a label is required for application to the file jacket.

4. Closing a volume automatically means the creation of a new one, incremented by one.

5. No file will ever exceed 999 volumes.

6. All volumes must be recorded in the application in order for the paper-dependant functions to operate.

7. Volumes are sometimes not closed on the last date of the correspondence within the volume. The application should therefore never automatically assume that the present date is the closing date. The system should allow an overwrite of the date closed field.

When working in volume maintenance, the information manager should first select the prefix and file number to be worked on. The prefix should be obtained from a scrollable list of valid prefixes or entered directly, if known. The file number field should invoke the look up function (see section 11) or should be entered directly, if known. From this point, the information manager should be able to accept the selected prefix and file number and proceed with volume maintenance or clear the fields to begin again, or exit the function.

Once both the prefix and file number have been selected and validated, the volume information should be presented as a scrollable list and should include the sequential volume numbers, from and to dates of the volumes, their location, security levels, and the open and closed status of the volumes. From this list, the information manager should be able to insert a new volume. When entering information on a new volume, the same fields as listed above should be presented. The volume number should be defaulted to show the current or highest volume number for that file. The TO field should automatically display the present date and the remaining fields should appear with the same information as for the previous volumes. It is very important that the information manager be able to overwrite all fields because the default information provides only a guide as to what information should be entered for the creation of the new volume. The information manager should also be able to delete a volume from this location. Deletion should always be accompanied by a confirmation message. Prior to allowing the deletion of a volume, its status should be closed. The information manager should also be able to edit volume information at this point or exit the function and return to the prefix/file number selection fields. It should be noted that this component should be very structured, yet flexible. All fields should be editable by the information manager but all should be mandatory.

8.0 ESSENTIAL RECORDS

Essential records are governed by federal legislation and government policy and therefore must be accommodated by the CIMA. The Emergencies Act, in effect replaced the War Measures Act, and is broad legislation which enables the federal government to fulfil its constitutional responsibility to provide for the safety and security of Canadians during national emergencies. The Emergency Preparedness Act sets out the role of Emergency Preparedness Canada and provides for effective civil emergency preparedness and for cooperation between federal, provincial and municipal governments in this area. The "Management of Government Information Holdings" policy includes direction on the management of essential records. In the section on Policy Requirements, sub-section Maintenance and Protection, the policy states that "Government institutions must: identify and protect essential information holdings (as defined in the Guide to the Preservation of Essential Records published by Emergency Preparedness Canada)."

In general, the essential records of any institution should comprise no more than 1 to 2 % of the total volume of records of the institution. It is important to stress the difference between records that are "desirable" as opposed to "essential" in the selection of the records. The definitions must be strictly adhered to if the volume of essential records is to be kept to a minimum. Also important to the program is the requirement for the selected records to be complete so that a person who is relatively unfamiliar with the topic would be able to function effectively. This often means summarizing or changing the form of the original record to facilitate its use. The Essential Records Program itself should be kept as simple and economical as possible since it must be updated and kept current on a regular basis.

The application should be able to classify and distinguish between the three categories of essential records. A field in the document profile, for use by the Information Manager only, should be available to indicate whether or not a document is essential and if it is essential, in which of the three following categories.

Category 1 Category one records comprise those records essential for a government, operating from fallout protected emergency facilities, to govern responsibly from the onset of the nuclear "attack period" until the shelter period is over and radiation levels have fallen to safe limits.

Category 2 Records in category two comprise those records considered essential to re-establish the organization, functions and responsibilities of government. They are required at a time during the "survival phase" when radiation levels have fallen to safe limits and the need to occupy fallout protected emergency facilities has passed.

Category 3 Category three records comprise those records that are essential to re-establish the basic rights of individuals and corporate bodies. They will be required during the "recovery phase", that phase immediately following the "survival phase".

Records that do not fit into one of these three categories should be considered as non-essential.

The CIMA should provide extra protection for the records flagged as essential due to their added importance. The system should maintain a profile field marked essential, category 1, 2, 3 or non-essential, for each document and paper file. The system should also provide a convenient means for the information manager responsible for the management of the essential records, to flag specific documents, individual files or ranges of files as essential. Only the information manager should be allowed to change the status from essential to non-essential or vice versa to ensure program integrity. The storage of essential records is also unique. Essential records are to be stored off-site in a secure location, as dictated by government policy.

At least three media alternatives are available for the storage of records deemed essential, namely electronic, micrographic and paper. A fully equipped office system which follows the specifications outlined in these functional requirements should be able to accommodate all three media types, allowing flexibility in the choice of media best suited to each category of essential records.

The off-site storage of essential records is, naturally, of utmost importance. The National Archives of Canada maintains a nation-wide network of federal records centres with facilities to ensure the appropriate protection of dormant records of varying levels of sensitivity. The Archives also maintains secure sites for the remote storage of a set of the vital or essential records of each federal institutions. These sites meet security standards and have environmentally controlled vaults for specialized storage requirements of microfilm and EDP records. Temperature and humidity are maintained at constant levels and are controlled through monitors and physical inspections. Each government department, agency and crown corporation must store a set of its essential records at these sites.

The CIMA should posses a function called PROCESS ESSENTIAL

DOCUMENTS which would allow the Information Manager to specify, for documents with an essential status, either to print to paper or copy to an electronic storage unit for off-site storage purposes.

9.0 SECURITY CLASSIFICATION

Security concerns appear throughout the functional requirement as they are an integral part of many separate functions. Some of the key security issues are highlighted in this section but also appear in the corresponding subject areas of the report.

It is necessary that, with few exceptions, all users within an institution should be able to view all file **titles** within that organization. The system should not restrict file title access, but should restrict access to the **contents** of files. The system should display a complete list of prefixes and their descriptions. Access to prefixes to which a user does not have access should be denied. In the case that a title may be security classified, an alternate non-classified title should be developed and used in its place.

The CIMA should allow up to 25 security levels to be defined by individual institutions rather that the standard five to allow sufficient flexibility to accommodate unique security requirements or policy changes.

The security level should be defaulted to unclassified, given that most information produced falls into this category. Security level should be configurable, however, to accommodate situations where a user might be working solely with classified or protected information.

Paper documents that bear a security classification of Protected or higher and designated for destruction must be physically destroyed at destruction time according to the Security Policy of the Government of Canada. In the case of electronic documents, the CIMA should allow the Information Manager to segregate security classified documents by security level thus facilitating the destruction process.

The storage media *must be erasable*. The proper treatment and handling of security-classified documents is extremely complex and made even more problematic with a non-erasable media. Also, there are uncertainties surrounding the legal implications of using non-erasable media for classified or designated information. In addition to these two points, it is likely that erasable technology will become more prevalent in the next few years and will therefore not be a problem.

A series of labels, each sequentially numbered, should be printed for application to each electronic storage unit. Each label should show the highest security level for documents on the electronic storage unit.

10.0 MULTI-LINGUAL SUPPORT

This section deals with the approach to multi-lingual support for the application. The user need not be aware of the functionality of this aspect of the system, but should be made aware of the options it will provide.

In order to comply with federal policy and legislation governing the use of both official languages, the CIMA must provide all users with the option of working in either official language. All on-screen field labels and help messages should be made available in both official languages. The application manager, information manager and the user should be able to select the language of choice and the system should be able to default to that choice at login time. Multi-lingual support could be provided by hard coding messages in files compiled with the program code.

11.0 LOOK UP FUNCTION

The look up feature should provide access to an on-line subject index of corporate information holdings. It is recommended that the structure mirror the paper system or vice versa to ensure consistency of information indexing at the corporate level (see section 3.0 System Architecture and Configuration, Filing System Organization). This portion of the functional requirements involves the document classification process, a process which has been well established in the traditional information management domain. The user looks up or selects the appropriate file number for the document being filed based on the subject contents of that document. It is this process that the look up feature addresses.

The look up feature should allow the user to execute the following functions: view a special field of explanatory text about the primary subject. Ideally, this description should be seen only if the user desires. The user should also be able to query for subject titles using key words, be able to view the full text of the description and the corresponding file number, be able to easily move to the previous and next hit, select an entry for viewing, or abort the query. This feature is used at filing time, search and retrieval time, at the time of verification and ideally, outside the application at any time the user desires. It should be a dedicated function available from a main menu and should be made available from within any application.

The look up feature should be modelled after methods currently being used in the paper world. At present, the user moves through the hierarchy of the subject index by scanning subjects, sub-subjects and file titles. Titles are more important than the numbers which represent them in most cases. The look up feature should simulate the mental processes of the user, that is, it must simulate this hierarchical subject review process. The look up feature should therefore mimic the information classification index where a user moves from the top or general to the bottom or specific levels of the index hierarchy. The following diagram helps to illustrate this idea.

LEVEL	FILE NUMBER	SUBJECT TITLE
Prefix	AD	Administration
Section	3000-3999	Property Management
Subject	3020	Accommodation
File	10	Moving Schedules
Sub-file	3	Operations Branch
Sub-sub-file	6	Move to Research Complex
Sub-sub-sub-file	2	Computer System

All titles within the specific subject hierarchy must be made visible in order to provide context. If all titles are not visible to the user, the subjects become ambiguous. For example, the title "computer system" without the context of the previous titles, could be interpreted in a variety of ways.

There should be a minimum of three ways to look up information on subjects within the on-line information classification index: navigation, focused navigation and querying. Each of these is detailed below.

11.1 NAVIGATION

Navigation involves the ability to browse through the index at will in a logical, ordered manner. The index should be presented to the user in the form of a scrollable list of prefixes from which the user may select one. A list of section titles within that prefix, sorted by their order of occurrence, that is, numerical order should then be presented . The user should be able to browse up and down through the list of titles and select the desired one. This selection indicates that the user would like more details on this section, that is, that the user would like to move down a level in the hierarchy from this section title. While still displaying the section title, a scrollable list of subject titles which directly relate to the selected section should appear. These steps should be repeated to the lowest available level, always displaying the titles of all previous levels selected to provide context.

11.2 FOCUSED NAVIGATION

Focused navigation involves the query function. The user must be able to query the system using key words or word combinations that the user would like to find in subject titles of interest. The same Boolean logic must be available as in other search features, that is, and, or, but not... options. The query results should be displayed as a result list where the user can navigate through the list of hits to select the appropriate one. The user should also be able to see the "neighbours" of a hit, that is, their the preceding and subsequent titles. By selecting a title from the result list, the user should be able to progress down a level to view the sub-subject titles related to the parent title. The ability to repeat this navigation process until the desired subject is found should be made available. Once the desired subject is located, the user should be able to select it and have the corresponding number displayed.

11.3 QUERYING SUBJECT TEXT

The third look up technique should allow for querying using key words and search strings. The search strings or word combinations must be marked, usually by placing them in single or double quotation marks. The same Boolean logic option should be available as above. Also, the user should be able to search on truncated words. Three examples of this option are included below.

Search on GEN* will result in GENeral
GENeration
GENerations

Search on *TION will locate automaTION
locaTION

Search on ?ill* will locate fILLing
bILLiard

There should be a limit imposed on the number of hits obtained in the result list. The number should be determined by specific user needs. If a search yields more than the pre-determined maximum, then a message should be issued stating that the search is too general and should be re-defined or re-focused. The system should provide the user with the option to return to re-enter query information or abort the query altogether. A second option would be to truncate the list at the pre-set maximum and provide the user with a message to this effect.

11.4 THE RESULT LIST

The result list should consist of each subject title that met the specified criteria in the query. The titles should be sorted by levels, from top to bottom, following the hierarchy of the corporate index. Within each level, hits should be sorted numerically. The user should be able to scroll up and down the list at will. The user should therefore be able to switch to navigation mode if desired.

12.0 SEARCH AND RETRIEVAL

12.1 THE SEARCH FEATURE

The most common type of search should be executed via a document content search screen, filled out by the user. The capability to search for any document should exist, whether the document is in electronic form or not, whether it is stored on primary or secondary storage. For non-electronic documents, a queryable free form text field which denotes a verbal summary of the document or a list of key words for that document should exist. Ideally, for electronic documents, the full text of the document should be queryable. If this is not possible, the document summary must be queryable. The search must include the full OR, AND, BUT NOT... options inherent in Boolean logic. Combinations of words or "search strings" must also be allowed. Search results should be displayed in the form of a listing.

More refined or expanded searches must also be possible. A document profile search should include querying ability on the following fields: document number, version number, date filed, filed by, to, from, type, subject, security, and essential. In addition, specific query features at this level designated as being accessible only by the application manager should include the ability to search on the cross-reference field for related information, the capability of searching on documents that have yet to be verified, and the ability to search by medium. These last three options are unique functions of the information manager and would not be required by the average user.

Another search option that should be available is a file profile search. This type of search is the most restrictive, is least likely to be used and involves a search at the file level. The following fields should be available for query purposes: prefix, start file number, end file number. In this scenario, only one prefix is searched at a time, the default being the user's home prefix. Ideally the user should be presented with a list of prefixes from which one can be selected. If no 'end file number' is entered, then the search will include all those numbers below the selected file number in the hierarchy.

The results of all searches should yield a result list.

12.2 THE RESULT LIST

The result list should be available in two modes, to be configured by the application manager. The first list type should display only those hits that are accessible by the user. No indication of restricted hits should be presented to the user in this case. The second option would be to present the user with a list of the total hits with an indication of which are accessible and which are restricted. The option is determined by management style and preference.

The result list should first be presented as a synopsis of each of the documents that meet the specified criteria of the search. For each of these documents, the following information should be displayed: document number, date created, subject and availability. All entries should be labelled with a sequence number to denote its order in the result list. Entries should be sorted in document number order. Ideally, the subject should be a one-line entry but should allow for horizontal scrolling for viewing of the full subject where necessary. Only current versions of the documents should be displayed with the option of expanding the search to include previous versions if desired.

A default should be established for the maximum number of documents to be included in any result list. Upon reaching this threshold, the user should be warned via a message and given the option to continue or abort the search. The user would then have the option to re-define the search and execute the query again, or continue. From the result list, the user should be presented with the following options: select an entry and display its profile, or return to the search screen and modify the search parameters. If an entry is selected, the following options should be available: view profile, view document, print document profile and text, retrieve the document into user's personal workspace, view next and/or previous profiles in the result list. If "view profile" is selected, the following items should be included in the document profile: document number, version number, date filed, filed by, to, from, type, subject, security, and essential. In the case of view document, print and retrieve, these options constitute access to the document and the action must be recorded in the audit trail. (See sections 14 and 17, Access Control and Audit Trails)

Whenever a user selects a document from the result list, a field called "Paper Documents" should be displayed and, if selected, the CIMA should display a corresponding list of paper volumes, their status, and start and end dates for the file recorded in the selected electronic document. This would allow the user to get a sense of what related paper documents were associated with the selected electronic document.

A user should be able to amend the result list at will by deleting specific documents or ranges of documents. Also, the result list must be formatted as a report with a print option for user reference.

More advanced search functions should be made available by the adoption of full-text search features.

12.3 RETRIEVE FEATURE

The retrieve feature retrieves a specified document from the corporate filing system and copies it from the corporate domain into the user's personal domain. The audit trail should be updated to reflect this access action. (See sections 14 and 17, Access Control and Audit Trails) There should be two retrieval options available to the average user. One is a dedicated function where retrieval is accomplished by document number. If the user knows the document number, the system should retrieve the latest version of that document and provide the user with the option to alter the version number if desired prior to copying the document down to the user's personal domain. The second method of retrieval is a sub-function of the search process. Using the search function described above, the user should be able to locate the document and version desired and use a retrieve option. The user should also be able to request that the Information Manager retrieve any document that has been migrated for non-disposition purposes.

All documents retrieved should be copied into a separate, dedicated directory (user's default directory) that should be specified when the user is created in the system. If the document already exists in the user space, a prompt to overwrite should appear. If no is selected by the user, then the retrieve operation should be cancelled or an option to re-name the existing document should be provided.

Whenever a document is retrieved by a user, an audit trail entry must be made to record the access to that particular document. (See Chapters 14 and 17, Access Control and Audit Trails)

13.0 DOCUMENT SUBMISSION

The submission of a document refers to the process of filing a document into the official corporate filing system. The system should be able to handle both the filing of electronic and non-electronic documents but should only allow the filing of new documents. The re-filing of documents that have already been filed previously should be carried out through Personal Document Management, section 6.

13.1 VOLUNTARY VERSUS INVOLUNTARY FILING

There are two alternative means of capturing electronic information in the corporate system, namely voluntary and involuntary collection. In the involuntary mode, the application would capture all documentation created in the user's personal workspace and retain a copy of this documentation in the corporate domain. This is not seen as a viable alternative at the present time. Users would likely view this kind of mandatory option as an invasion of privacy. Therefore, these functional requirements will deal only with voluntary collection where the user decides what should be included in the corporate system and when it should be captured. This alternative is much closer to the rules of the paper system and would likely be acceptable to most users. On the other hand, the voluntary collection alternative requires highly trained and motivated users to ensure that the goal of preserving the organization's corporate memory is adequately being met.

13.2 THE SUBMISSION MODEL

Submission of a document should begin when a user has completed the process of preparing an electronic document using an application package, usually a word processing application. There exists an underlying assumption that the electronic document exists on a disk service available to the workstation, that is, the document is reachable through a specified DOS path.

As a first step, the system should require the user to identify the document as an electronic or non-electronic document, for the handling processes associated with each of these two formats will likely be different. In either case however, the user would be required to complete a document profile. The document profile is a collection of information or data fields, recorded by the CIMA, to assist in the management of the document. The profile should be similar in concept to the information usually found in the header of a typical memorandum. The document profile fields should include prefix, document name and path, file number, to, from, security level, document type, subject of the document, a short summary of the document, and if required, a list of those users permitted to access the document (refer to Access Control, section 14). All of these fields should be editable by the user. If the document to be submitted is electronic, then the document name and path, and document type fields should be mandatory. If the document to be submitted is an non-electronic document, then the summary should be mandatory for retrieval purposes. The file number should be entered by the user. The user must therefore be able to quickly look up the appropriate file number in an on-line, readily available corporate index (see section 11). Several fields should automatically be completed by the system, namely document number, version number, date the document was filed, who it was filed by, and the date the document was last accessed. These fields should be displayed to the user for information purposes only and should not be editable by the user.

Four other fields should be associated with the document profile. They are cross-reference numbers, verified, linked, and essential. These four fields should not be displayed for the average user. They should be restricted to the Information Manager responsible for the integrity of the information contained in the corporate system. An entirely separate process, to be used exclusively by the information manager, should be established to enter and modify these fields.

Of the fields available for user viewing, fixed list selection should be offered for prefix, file number, to, from, security level, and document type. In the case of file number, the fixed list selection should automatically invoke the look up function. (See section 11)

13.3 FIELD DESCRIPTIONS AND FUNCTIONALITY

The prefix field should always be defaulted to the user's home prefix since the user will be carrying out most of the filing within that prefix. The document name and path field should be configured to default to each user's system path name. The file number field should be defaulted to 9999-99-9 as an example of the number structure to be entered in this field.

Upon entry to both the TO and FROM fields, a scrollable list of system users should appear in alphabetical order and the user should be able to simply select a name from the list. The user must also be able to override this fixed list and enter original text to accommodate names that are not on the list. The TO field should be configurable but should default to TO FILE. The FROM field should default to the user's name and should also contain an override feature to allow original text to be entered in this field. The option to change the default name should also be made available to accommodate users who compose documents on behalf of others. In summary, the user should be able to enter information in the TO and FROM fields without major restriction. As a safeguard against someone entering another user's name instead of their own in the FROM field, the system should automatically record the real sender's login name in the document profile in a field called FILED BY.

The security level should be defaulted to unclassified, given that most information produced falls into this category. Security level should be configurable, however, to accommodate situations where a user might be working solely with classified or protected information.

The system should maintain a list of all possible document types used within the organization. This list should be configurable to accommodate changing needs and each document type should be associated with a unique code or identifier, a three digit sequential number is one possibility. Then, when the user invokes the fixed list for document type, the system should present a list with associated identifiers. When a document type is selected, the system should automatically insert the corresponding code into the database. This will avoid the user being required to memorize cryptic mnemonics such as WP for WordPerfect, WS for Wordstar, HA for Harvard Graphics, etc. The document type field should default to the most common document type created in the organization, in most cases, the most current version of WordPerfect.

13.4 PROCESSES

The SUBMIT function should permit the user to file a document with its profile, into the corporate domain. The system should notify the user of the document number assigned by the system and should offer the choice of retaining a copy of the document in the user's personal workspace through a prompt. If the user elects to retain a copy of the document, it should be stored in the user's retrieve directory. The system should be conducting a verification to ensure that the document has not been filed previously by using the document number and version number associated with the document.

The QUICK VIEW function should allow the user to view a rendition of the document without having to load the original application. This will facilitate the user's determination of a file number if the exact subject is not fresh in the user's mind.

The user should also be permitted to change from electronic to non-electronic document submission without having to exit this portion of the software. Users should be allowed to exit the submission process at any time.

14.0 ACCESS CONTROL

Access control refers to the concept of regulating access to certain functions within the system based on a user's need to access particular functions. The application should implement an access control layer between the users and the system itself. All requests for information should be filtered through this layer to prevent users from gaining access to information without the appropriate level of authorization. The system should allow the application manager and the information manager varying degrees of authority within the application. Access control requires that all users, including privileged users, be granted access rights to all or portions of the system. For example, a user may be able to access one or multiple prefixes, or may be restricted to only one small range of numbers.

In some organizations, it may be acceptable to allow users who do not have access to certain file blocks to see the file titles to which they do not have access.. In other organizations this would be deemed unacceptable. The system should be configurable to meet these access control criteria.

14.1 DOCUMENT ACCESS LIST

If all members of an organization are expected to use an application for the management of electronic information as described by these functional requirements, there must be a method by which documents can be protected from random access, aside from security segregation. For example, management is often working with information of a sensitive nature that is not security classified or designated. Topics such as future management strategies, downsizing exercises, budget cuts or personnel problems, are all a regular part of managing an organization in the federal sector. None of these topics are necessarily security classified but all may need to be restricted on a need-to-know basis. In order to address this problem, the system should allow documents to be restricted to selected individuals within an organization. This function should be similar to the management of such information in the paper-based environment, where access to particular files is restricted to lists of personnel based on the need-to-know principle.

When a user submits a document to the corporate domain, the system should allow for the creation of a document access list as part of the completion of the document profile. If granted the proper authority, the user should be able to create a list of users that are able to access that particular document. If no list is created, all users should be able to access the document. It is extremely important that this function be controlled to prevent users from randomly restricting information in the corporate domain. The Application Manager should be allowed to configure this option on an individual user basis. For example, it may be a management decision to provide only senior managers with this capability. Sensitive information generated by senior managers could then be restricted from access by the average user.

In searching for information, users expect to locate all information. In order to satisfy this demand, the system should display the total number of occurrences that are accessible and also show how many occurrences are restricted. This would allow users to seek authority to view the restricted information if desired.

15.0 RETENTION AND DISPOSITION MANAGEMENT

A key element of any effective information management program involves the preservation of information designated as archival and the destruction of obsolete information, according to approved retention and disposition schedules. Current paper-based methods of retention and disposition scheduling must be adapted to the electronic environment. This section of the functional requirements describes the processes currently in place for retention and disposition management and outlines the required transition of traditional retention and disposition methods from paper-based application to the electronic environment. The requirements proposed in sections 15 and 16 of this report have been named Electronic File Migration or **EFM.** The capabilities outlined herein represent an essential component of a system for the management of electronic records.

In order to effectively understand this section of the report, the following three definitions of "document" must be understood.

1. **Electronic Documents.** Documents that are entirely in electronic form, and are stored within and managed by the CIMA. A profile of information about the document, such as subject, author, etc. is associated with it.

2. **Non-electronic documents.** These documents are recorded by the CIMA but are not stored in it. They are treated by the CIMA exactly the same as electronic documents. A profile of information about the document, such as subject, author, etc. is associated with it. Some examples might include a bound annual paper report, a VHS video cassette, a set of photographs.

3. **External Documents.** Documents entirely outside the CIMA. They are not recorded within the CIMA, nor does the CIMA record anything about them. They may be held in any form, i.e. electronic, paper, microfilm, etc. For the purposes of migration, particularly disposal to National Archives, external documents have to be associated with, or sent with, the electronic and non-electronic documents managed by the CIMA.

15.1 GENERAL PREMISES

The following premises have been used to develop the functional requirements surrounding retention and disposition management. All rules are based on accepted information management and archival principles.

1) At any point in the lifetime of an electronic document under the control of an Information Manager, it should be possible to readily render the document to paper.

2) The Information Manager should be able to add and revise retention and disposition information on any electronic document at any time.

3) The CIMA should allow an Information Manager to purge individual versions of a document, and entire documents as required.

15.2 FILE MIGRATION IN THE PRESENT SYSTEM

PREMISES

1) It is assumed that the Information Manager is able to migrate information to the National Archives at any time. The securing of the necessary background approvals in the retention and disposition process are not of concern to the CIMA. The CIMA should simply supply fields to record these background approvals.

2) At final disposition time, the National Archives requires a comprehensive "inventory report" to accompany the information sent for archival retention.

3) At any time, the CIMA must allow the extension of the retention period of any record under the Information Manager's control. However, there are certain rules and limitations to which s/he must adhere. These rules and limitations are detailed in the next section of this report, sub-section "SCHEDULE MAINTENANCE".

4) Paper documents that bear a security classification of Protected or higher and designated for destruction must be physically destroyed at destruction time according to the Security Policy of the Government of Canada. In the case of electronic documents, the CIMA should allow the Information Manager to segregate security classified documents by security level thus facilitating the destruction process.

5) Unscheduled documents exist indefinitely and are in the realm of responsibility of the Information Manager, in either an active, semi-active, or dormant state until such time as an approved retention and disposition schedule is created. However, the Information Manager should be able to arbitrarily designate documents for transfer to the National Archives as required. This process is known as "direct transfer". This is a very uncommon occurrence. One of the reasons

direct transfer may occur is that old documents for which a proper schedule had never been approved may be re-evaluated as being historically significant. This direct transfer would require a special, one-time authority number. The CIMA should allow for this equivalent capability. In the present paper system, the Information Manager may arbitrarily designate individual documents, series of files, single files, single documents, or any combination for direct transfer to the National Archives.

6) An agency is never obliged in any way to transfer dormant documents to an outside storage facility, such as the Federal Records Centre network or commercial storage facilities.

PACKAGING FOR SHIPMENT

When applying retention and disposition schedules to paper documents in the present system, the documents are packaged for shipment. The rules of the packaging are as follows.

1) Files contained in each box must be in some logical order. The order is left to the discretion of the originating agency, and is usually in accordance with the agency's subject classification system, usually file number order.

2) Volumes must be in numerical sequence within each file.

3) The documents within each volume must be in proper order. The order is determined by the originating agency, but usually, approximately 80% of the time, this translates to simple chronological order. It should be noted that is the case for paper documents only. Information held on other media, electronic records or microform for example, is not necessarily treated in this manner at present.

The Information Manager can not act without an *authority number*. This authority number is a form of proof that the schedule has been approved by the National Archivist. There is one instance where an Information Manager may enter scheduling data for newly created files. Where the new file is a *designated* case or linked file, the new file inherits the authority number, and scheduling data from a parent or adjacent file.

At the time of file creation, most files do not have known scheduling data. It is expected that there will be a gradual increase of the percentage of records with approved schedules. It is expected that over the next 2-5 years, there will be a shift such that a greater percentage of files will have known scheduling data at initial creation time.

Migration in this report has been defined as the physical movement of documents. There are two types of migration, namely *disposition* migration and *non-disposition migration*. Disposition migration involves migration for the purpose of transfer of ownership. Non-disposition migration involves the migration of information for any other purpose, such as to save physical storage space.

Once a document's activity level declines, it may be considered "dormant". If it has an approved retention and disposition schedule, it becomes dormant according to the period specified in the schedule (subject to change by the Information Manager based on several rules and guidelines). If the document does not have an approved schedule, most agencies will apply a general rule of thumb to determine how long to keep information in active storage areas. This rule of thumb is entirely determined by practicality within the agency. In the case where there is no approved schedule, it is illegal to destroy information. Therefore, the information may be considered to be dormant for an indefinite period of time.

Once information has become dormant, it may be sent to an off-site storage facility or it may be retained within the agency. Finally, at the end of its term, that is after expiry of the sum of its active and dormant retention periods, it will be either be destroyed, or sent to the National Archives for evaluation.

MIGRATION TO AN OFF-SITE STORAGE FACILITY

There are several assumptions regarding the migration of information to an off-site storage facility imbedded in these functional requirements.

1) There are no restrictions or limitations as to what the off-site storage facility will accept from a client.

2) The Information Manager of an organization is not obligated in any way to send documents to an off-site storage facility, even if the documents are in a dormant state.

3) The off-site storage facility will accept documents with retention periods of any length.

4) The off-site storage facility routinely references individual files or, in the case of media other than paper, information entities, from containers. The off-site storage facility will never remove individual documents from files or reference information from within an entity.

In other words, the off-site storage facility will never retrieve any unit of information smaller than a file or entity. The off-site storage facility will likely maintain retrieval codes for files or information entities, but not for individual documents that may be contained on an individual file. The same principle applies to media other than paper.

5) The Historical Resources Branch of the National Archives recommends that for archival purposes, electronic media not be re-used. These media should only be used once to record historical data.

6) The off-site storage facility will not assume responsibility for media formats that fail with time.

MIGRATION TO THE NATIONAL ARCHIVES OF CANADA

Once the disposition date of information has been met, it may be transferred out of the control of the Information Manager to the control of the National Archives of Canada. The Information Manager may have documents that have met their retention period, but for internal reasons, the Information Manager is allowed to extend the retention date. When s/he is satisfied that the documents are ready for transfer, an inventory listing and a history report must be prepared by the Information Manager.

Once this is completed, the Information Manager then contacts National Archives officials, who are obligated to accept the transfer. If there is a delay in the process for any reason, the Information Manager merely postpones the transfer until a later date. Upon transfer of the information, the National Archives ensures that the inventory listing is satisfactory and that indeed the information has met the approved schedule.

Once the documents have been transferred, the National Archives corresponds with the originating agency to confirm that they have accepted the documents. At this point, the National Archives notifies the agency of any destruction it is planning to carry out (60 days notice). If the originating agency approves the destruction, the National Archives will proceed to destroy the information.

RETENTION AND DISPOSITION STATISTICS

The following statistics should be kept in mind when designing the Electronic File Migration section of the CIMA.

1) At the time of file creation, less than 5% of all operational files have known retention and disposition information.

2) Approximately 1/3 of all federal government files are administrative. These files have a blanket, approved schedule in the form of the General Records Disposal Schedules of the Government of Canada (GRDS). Therefore, Information Managers will always know the scheduling data for this particular information at file creation time.

3) Approval for extensions to scheduled files takes place much sooner than for new files with no current scheduling precedents.

VOLUME "STATES"

A file's lifetime could be considered to be its total retention term, or the sum of the active and dormant retention periods as specified earlier. During its lifetime, the paper volumes comprising the file will each pass through up to three "states", active, semi-active, and dormant.

A volume is considered *active* if the active retention period has not yet elapsed, relative to the most recent entry in the file. If, for example, the file has an active retention period of five years, one of its volumes leave the active state automatically according to one or more of the following conditions, or at the discretion of the Information Manager.

> The most recent document entry is at least five years old, i.e. nothing has been added to the volume for the past five years.
>
> The case is now closed (terminated).
>
> The volume is no longer used (low level of access).

A volume enters the <u>dormant</u> state after the active retention period has expired. It is migrated either to an organization's own dormant storage facility, or to an off-site storage facility. This physical relocation is based on the assumption that dormant files rarely need to be accessed, and a longer access time would therefore be acceptable to users. By the time a volume becomes dormant, the access activity has usually slowed considerably. In some cases however, the Information Manager may override the usual procedure and retain the volume in active space for operational reasons.

It is estimated that approximately 30% of dormant volumes are sent to off-site storage facilities. This fact will have serious implications for the CIMA because it means that fully one third of the documents a user may be looking for may not be available. Two issues arise from this fact.

1) There can be no reduction in the quantity or completeness of data stored about a document, if stored at a secondary location. If the documents are located in the off-site storage facility, they are still considered to be under the organization's control and must be retrievable.

   It is estimated that the quantity of electronic documents stored at the off-site storage facility will initially be quiet high, but will gradually diminish as the originating agencies acquire technology and experience in storing and retrieving large quantities of electronic documents.

2) If up to 30% of the CIMA's documents are stored off-site, it is presumed that they may be recalled at any time. The CIMA must therefore keep track of which documents are off-site, and must be able to quickly and easily re-integrate them back into active storage. If the documents remain in the off-site storage facility through to final disposition, the Information Manager must update the CIMA so that information on the transfer of these documents is present in the system.

There are two possible ways in which a document may be added to a dormant volume. If the incoming document is "current", a new (current) volume is created to store the document. If the incoming document has a date which falls within the range of one of the dormant volumes, it is added to the appropriate volume and the dormant volume's state will remain unchanged.

A file can be made up of one or more active closed volumes. Some or all of these volumes may be removed to the organization's dormant storage facility in order to save space in the active area. Such volumes are still considered active, but usually have a lower activity rate. They are still technically active, but there is very little access to them. If a closed volume has aged to the point of final disposition, it may be destroyed, assuming this is its prescribed final disposition, even though more recent volumes of the file may still be current.

A file does not necessarily migrate from state to state in its entirety. Individual volumes may migrate from state to state independently of the rest of the file. For instance, a file's current volume may be considered active, and one or more of its closed volumes may be dormant, assuming the active retention period has expired.

The CIMA must be aware of the particular state of documents at all times because the rules and processes for handling the document change when its state changes.

In the paper system, the existence of volume states is loosely based on two criteria. The first involves physical criteria. Files are physically stored in three different locations, depending on the state (active, semi-active, or dormant), so that physical storage space can be managed effectively. The second is based on access criteria. The users of the system, and the Information Manager, are willing to accept an increase in the time required to access files in the semi-active and dormant states, because the frequency and urgency of access declines with time.

CASE OR LINKED FILE HANDLING

Of all federal government departments, about 20% of them create predominantly case or linked files. Canada Employment and Immigration and Revenue Canada Taxation are two such departments. The remaining 80% of departments create mostly subject files, but an average of 15% or more of their holdings are linked files.

A "case file" is one that deals with a specific subject, individual, event, or linked series of transactions. This report will refer to case files as "linked files". This contrasts with the typical subject file, which is based on the related subject matter contained within it. Like a subject file, a linked file, such as a file on a contract, has ongoing activity of some level. The CIMA must concern itself with linked files because it must be able to handle the unique manner in which their disposition dates are determined, i.e. relative to a termination/close date of the file. Some examples of final disposition dates for linked files may be the end of construction of a building, or the end of a contract. A file may be designated a linked file under any combination of the following circumstances.

1) The end of the subject's activity period is indefinite, or not known. For example, if an inquiry is set up, documents may continually go into the file, but the Information Manager will have no idea when the inquiry will end.

2) The activity level of the file is unpredictable, or steady throughout its lifetime. An example might be a file on a particular person, or a file on a particular airplane. While the person/aircraft still exists, there could be unpredictably heavy periods of activity. For example, after a period of low or no activity, suddenly the person could become quite important, or the aircraft could crash and be the subject of an investigation.

Basically, there are two key reasons why the CIMA must distinguish linked files from subject files. First, this distinction is necessary for reporting purposes. The Information Manager may wish to generate reports of linked files to show activity levels, migration status and so forth. The Information Manager may periodically want to view these files separately from ordinary subject files. Second, when migrating information to the National Archives, unlike subject files, a linked file must always have all of its volumes migrated at the same time, constituting a migration of the entire file. It should be noted that this is not the case when migrating information within the agency.

If the activity level of a linked file slows to the point where the Information Manager can confidently migrate the file, the possibility that access could suddenly be required exists and the file would have to be migrated back to the CIMA. It is therefore recommended that the CIMA migrate linked files in their entirety when being migrated to the National Archives.

In order to handle linked files properly, the CIMA should provide the following functionality.

The CIMA must allow the Information Manager to arbitrarily designate files as linked files, and to reverse the designation.

The CIMA must provide the Information Manager with a means of identifying and reviewing linked files for content and activity level, so s/he can make migration decisions.

The CIMA must force migration of all a linked file's volumes at the same time, when migrating to the National Archives.

16.0 ELECTRONIC FILE MIGRATION

This section describes the features of a CIMA that would be required to migrate, that is, physically move electronic records to and from a host system. The cornerstone of the Electronic File Migration (EFM) portion of the requirements are the current paper-based retention and disposition management practices. Many of these requirements are legislated or stem from federal policy and therefore must be incorporated in any such system.

16.1 ACTIVITY MEASUREMENT OF ELECTRONIC RECORDS

In order to judge whether or not information should be migrated, the Information Manager must be able to determine whether or not the information is being accessed by users. The Information Manager must therefore be able to query the CIMA to determine the activity level of selected documents. The system should allow for the querying of documents within a particular date range by a specified single file number, and a range of file numbers, as well as allow for the exclusion of certain files within a set of specified numbers. This will allow the Information Manager sufficient flexibility to search for the desired results.

The CIMA should offer the results of the activity measurement search to the screen or the printer. The results should include the number of accesses per electronic document and the number of accesses per corresponding paper file volume in all cases. The CIMA should already be tracking paper volume access through the charge-out and reservation functions. The Information Manager should be presented with three options as to how the results are presented. The CIMA should make the Information Manager aware of how many total accesses were made to the electronic document and the corresponding paper volume since their creation. The Information Manager should also be able to assess the number of accesses against time, that is, the system should provide the dates of access for the total files selected in the query, from the most recent to the oldest. The option should also exist to detail the total number of accesses by file number, including the date of the last access. These three permutations of the query results will allow the Information Manager to make well-informed decisions on whether information is ready to be migrated or not. In all three options, the query criteria used to generate the results should be presented to ensure that the operator is aware of the specified query criteria.

16.2 SCHEDULE MAINTENANCE

A key cornerstone of the CIMA is the input, editing and deletion of retention and disposition information, also known as schedule maintenance. The schedules should be *defined* at the file level, but *applied* at the volume level. As in the paper environment, each individual file title should be assigned its own independent retention and disposition data. There can be no assumed inheritance of retention and disposition data from related files. If a file has been designated as a linked or case file, the CIMA should be able to automatically transfer the retention and disposition data of the related files to the newly created one.

A schedule entry in the CIMA should consist of five components, each with a specific function. The active retention period, dormant retention period, specific disposition action, authority number, and whether the file is linked or not, should be recorded for each file.

In the ideal situation, all information would be formally scheduled upon creation and be associated with an authority number. In reality, much information exists without approved retention and disposition schedules and no authority number. It is important that the CIMA control the retention and disposition of the information contained in it in accordance with approved schedules. It should not, however, inhibit the Information Manager from managing the information in practical and effective ways. The CIMA should be flexible enough to allow the migration of information from primary to secondary storage, whatever the secondary storage may be, if the need arises, without restriction. It is estimated that average retention period for federal government information is 10 to 15 years. The CIMA will require criteria to evaluate unscheduled documents for migration eligibility. The CIMA should therefore default to a total file term of 15 years, five active and ten dormant, for all unscheduled files. There would be no authority numbers for such unscheduled documents. The AUTHORITY NUMBER field must be allowed to be left blank, signifying an unscheduled file. Whenever such files are displayed on the screen by the CIMA, the fact that they are unscheduled should be made apparent.

There are specific rules governing retention and disposition management and the CIMA should enforce these rules when the Information Manager edits existing scheduling data. The dormant retention period should either be zero or greater than two years and any changes to the file term, that is the total length of time information must be retained, must be accompanied by a change to the authority number. Also, any change in either the active or dormant retention periods which result in a net *decrease* of the total must be accompanied by a change in authority number.

When changing an existing authority number, the new number should not be the same as the existing number. However, the number may be the same as another authority number in the system. It should be noted that the term "new" authority number does not necessarily equate to a unique number, not used before in the system.

Also, any change in either the active or dormant retention periods resulting in a net <u>increase</u> in the total must be accompanied by a change in authority number if the original total was less than or equal to 10 years, and the replacement total is 1.5 times the original total, or if the original total was greater than 10 years, and the replacement total is 1.25 times the original total. No new authority number should be required if either the active or dormant retention periods are changed but the total remains constant.

All changes to existing schedule data should be recorded in the form of a background audit trail. The exception to this would be in the event that the total file term does not change.

16.3 DEMARCATION OF ELECTRONIC RECORDS

There are several assumptions regarding the demarcation of electronic records imbedded in these functional requirements.

1) All documents in a given file will have the same retention and disposition data.

2) It is estimated that, at present, approximately 10% of all files contain information stored on more than one media format.

3) It is assumed that the CIMA will be the mechanism whereby the Information Manager will record the existence of all media formats on a specified topic.

4) It is presumed that, within a given file, a physical volume may have a large date span, a microfilm cartridge spanning a full year's worth of documents, for example. The Information Manager will likely want to migrate information held on all media pertaining to a specific subject for a particular period of time in one migration. However, some media formats, like the microfilm cartridge, are difficult to conveniently split into arbitrary date spans. It therefore makes sense to select start and end dates of the other media formats to match the start and end date of the microfilm cartridge. This would allow a single date span to cleanly cover all media formats selected for migration.

DEMARCATION

According to information management theory, and as expounded by the MGIH policy, information should be managed in a manner independent of format (media). It is common for organizations to create and retain information in a variety of media formats on a particular topic. A user should be able to locate and examine this information, regardless of the combination of media types. The time 'and energy required to identify, process and retrieve this information should be minimized. A request for information should not have to be broken into many individual portions because of the requirements demanded by different media variations. The Information Manager should therefore strive to minimize the handling differences between media types to realize an effective information management program.

In line with this objective, a "volume", which is a paper-specific term, implies a date span, covering a range of time during which documents of a particular topic were created, collected and stored. Within this same period of time, media other than paper were also being created, collected and stored on that same topic. These other media types should be treated as one series of documents along with the paper files. Thus, the CIMA will consider a "volume" to contain paper as well as electronic documents.
A discussion of the management practices associated with the management of traditional paper volumes is necessary to establish the parameters of Electronic File Migration and to propose a method of including the electronic documents in the paper-based methodology.

In the traditional paper records management world, *files* are made up of *volumes*, and volumes are comprised of *documents*. Each of these three entities must possess certain characteristics. Files must be subject oriented, be identified by a unique file number and possess associated retention and disposition information including the active and dormant retention periods, a disposition action, and an approved authority number. A volume must also be identified by a number. Volume status must also be identified as open or closed, and the start and end dates of the volume must be recorded. At the document level, the date, media type, and topic, among many others as defined by the CIMA, must be present.

As previously mentioned all documents within a given file will have the same retention and disposition data, regardless of media format. It is, after all, the information itself that is scheduled, not the medium on which it is stored. However, cases exist where different media that contain information on the same topic have different schedules because of differing physical considerations. It is estimated that this situation occurs less than 5% of the time. It is therefore not necessary for the CIMA to accommodate this exception. If this situation were to occur, each different media type could be assigned a different, but related, file number, and scheduled separately.

The following chart illustrates how a hypothetical file might be divided into volumes.

```
VOLUME #1

        START = 1900/01/01 END = 1910/01/01    STATUS = Closed

                Document            Media       Date
                =====================================
                ========
                document 1          Paper       1900/01/10
                document 2          Optical 1905/03/11
                document 3          Map         1905/03/20
                document 4          Paper       1906/04/09
                document 5          Microfilm   1909/07/13
                document 6          Paper       1910/11/12

VOLUME #2

        START = 1910/01/01 END = 1920/01/01    STATUS = Closed

                Document            Media       Date
                =====================================
                ========
                document 1          Paper       1910/01/10
                document 2          Optical 1913/03/06
                document 3          Map         1913/03/20
                document 4          Paper       1916/04/09
                document 5          Microfilm   1918/07/04
                document 6          Paper       1920/11/24

VOLUME #3

        START = 1920/01/01 END = 1930/01/01    STATUS = Open

                Document            Media       Date
                =====================================
                =========
                document 1          Optical 1920/01/10
                document 2          Optical 1921/03/08
                document 3          Map         1923/03/23
                document 4          Paper       1925/04/11
                document 5          Microfilm   1926/07/18
                document 6          Paper       1929/11/11
```

In contemplating this example, it becomes apparent that the CIMA must record a media format for every document. Here, each volume contains a mixture of media formats. Should the Information Manager decide to migrate volume 1 for instance, s/he will be dealing with four different media formats. The Information Manager would therefore want to select the volume start and end dates to best suit the media formats. The Information Manager would select the volume start and end dates based on the predominate media format within the date range s/he was interested in. For example, it may be necessary to transfer a large number of paper files for sheer volume reasons, but a very small amount of microfilm may also relate to the selected files. Therefore, the Information Manager would set the volume start and end dates to migrate the paper. As a second step, all other media types relating to the files and that fall within the selected date span, should be located and included in the shipment of information to be migrated.

The usual process for a request for information from a user is as follows. When a user approaches an Information Manager and requests information on a particular subject, the Information Manager will determine whether the request is for information on one or all media formats. The Information Manager will then locate all documents in all forms that fall within the desired date range.

It is assumed that the CIMA will be the tool used to record all the media formats, location of the information, and other pertinent details required for retrieval and management of the information. Thus, the CIMA must have a feature whereby the operator can query for all documents, regardless of media format, in order to provide the Information Manager with the information to physically locate it for the requesting user.

The CIMA must have a feature whereby the user may query for information by media format. For example, a user may need to know all holdings of a particular media type that fall under file HR1000-10-2. Another example of such a requirement may be to identify all information within the organization held on a particular media format.

PREMISES

1) Information Managers need to move the boundaries of the start and end dates of the documents to be migrated arbitrarily, as required.

2) In the case of a disposition migration, all media formats must be included in the shipment. This is not the case for a non-disposition migration.

3) The term "volume" will not be used to represent a unit of documents for migration. Rather, the term Document Sequence will be used. The term "volume" has traditionally been used to describe paper documents contained within a single file jacket. In the electronic domain, this idea is obsolete. There is no need to artificially impose "volumes" on electronic documents. Storage capacity should allow for the storage of vast amounts of information without restriction. However, there is a need to divide electronic document for migration purposes. The term "document sequence" is therefore introduced as an alternative to "volume" for electronic documents. A document sequence will refer to a collection of documents, represented by a start and an end date. The term "volume" will be assumed to refer to the paper file folders.

In the paper system, a volume can pass between three states, namely active, semi-active, and dormant. When migrating electronic documents under the control of a CIMA however, the state should not be of concern. Rather, electronic documents should either be considered present, or migrated, as defined below.

Present Present on primary storage and immediately available in their entirety. There should be no need to migrate documents from a secondary storage device to retrieve documents. It is possible that a document has been migrated back to primary storage from secondary storage, and is "present", although perhaps for a limited time.

Migrated Not stored on primary storage, that is, recorded on a secondary storage facility. This information has been migrated by the CIMA. Only information *about* the document is stored on the CIMA and available for searching. In order for a user to retrieve the document, it will have to be migrated back to primary storage. Generally not available immediately.

These definitions clearly indicate that the *availability*, not the state, of the document should be of primary concern for the CIMA. In fact, the availability of a given document is completely independent of the state, for several reasons. Migration, particularly non-disposition migration, can occur at any point in a document's lifecycle, without any regard for the theoretical "state" of the volumes containing the document(s) in question. Also, volumes move through the states of active, semi-active and dormant, not individual documents. The CIMA should have no fixed concept of an electronic volume. Therefore it is not logical to attempt to associate the fixed paper volumes with their associated fixed date spans to the document sequences of electronic documents with their variable date spans. Lastly, the administration of the CIMA could become technologically complex if all three states had to be handled by the system.

16.4 MIGRATION PRINCIPLES

Migration is defined as the actual movement of documents from one physical location to another location. The CIMA should assume that only document sequences will be migrated. That is, the CIMA should not be designed to migrate units of documents smaller (documents) or larger (files) than document sequences. In the case of a closed, linked file, the entire file must be migrated according to information management practices. However the CIMA should treat such a case as the migration of a series of document sequences.

In reality, individual documents are occasionally migrated but this occurs less than one per cent of the time. Therefore, the CIMA need not allow for this exception.

There are five key considerations that the CIMA should take into account for information that has been migrated.

1) *Search capability.* How a user can search for, locate, and retrieve documents that have been migrated.

2) *Document format.* Electronic documents are originally created in "native" format and can only be accessed through the use of the application software originally used to create them. At what point in the document lifecycle should the native format be abandoned?

3) *Media Characteristics.* What criteria should the Information Manager use to select a particular media format for storage of migrated documents?

4) *Electronic Storage Unit Numbering.* All electronic documents stored offline must be stored on electronic storage units of some kind. These units must be identified in a unique manner for retrieval purposes. An electronic storage unit numbering scheme will be required.

5) *Retrieval.* How will a migrated document be delivered to a requesting user?

GENERAL MIGRATION PREMISES

1) The Information Manager should be able to select any media format, regardless of the retention period assigned to the information to be stored on this media format. If the format's longevity, as recorded by the CIMA, expressed in years, is less than the total retention period of the information, the CIMA should issue a warning, but take no action to prevent it from being carried out. It should be presumed that the Information Manager will take the most appropriate action in the given situation.

2) Within a series of archival documents designated for migration, some may be omitted, for a number of reasons. It is estimated that less than five per cent of all present paper migrations contain such omissions. There are two ways in which documents may be missing. Documents of a certain media format may be omitted or documents falling within certain date spans may be omitted. In order to handle these exceptions, the CIMA should provide for an explanation field for omitted documents.

3) The CIMA should also record the type of migration being conducted, that is either a disposition migration to the National Archives, or a non-disposition migration within the organization. The CIMA should permit different actions, depending on the type of migration being executed. Since non-disposition migrations are internal to the organization, that is, the information remains in the control and area of responsibility of the organization, the CIMA need not record explanations for missing documents. However, in the case of a disposition migration, the information is transferred out of the control of the originating agency to the National Archives, thus, explanations for missing information should be recorded.

4) If a paper volume selected for migration is part of a linked file, then, in the case of a disposition migration, all volumes of that file must be migrated together. This does not apply to non-disposition migration.

5) It should be possible to migrate documents, in the case of non-disposition migration, regardless of their state, either active or dormant. In other words, in the case of a non-disposition migration, the age of the documents within the designated document sequence may fall within either the active or dormant retention periods. In the case of a disposition migration, all documents which fall within the designated document sequence should have met the file term, that is the total of the active and dormant retention periods.

6) Both electronic documents and non-electronic documents, as stored

in the CIMA, are considered equal for the purposes of migration, and should thus be handled in similar ways.

7) Files, and therefore by implication, documents, can have their final disposition changed from KEEP to DESTROY, or vice versa, at any time with the proper authority from the National Archives.

8) A given document under the control of the CIMA should only exist in one place at a time. In other words, multiple instances of a document should not be allowed. A document should reside in only one of primary storage, secondary storage, or disposition migrated electronic storage units.

16.5 THE EFM MODEL

This section of the report details the fundamental EFM model. There are three possible models. Each of the three will be discussed, and the best will be recommended for implementation. Supporting reasons for the selection have also been included to facilitate the reasons for the selection.

MODEL 1

Model 1 is the simplest of the three schemes. In this first model, non-disposition migration would not exist and therefore, the CIMA need not be concerned with migrating documents back to primary storage for final disposition migration purposes. This model is attractive in its simplicity but has been rejected because the Information Manager requires the ability to carry out non-disposition migrations as required. The non-disposition migration process is essential to the Information Manager and will be required to respond to situations that will likely surface before a disposition migration is possible. For example, a non-disposition migration would be required to free up disk space or to reduce the quantity of documents in primary storage to facilitate system management.

MODEL 2

In the second model, non-disposition migration would be carried out to off-line electronic storage units. The subsequent final disposition migration would be accomplished by extracting documents directly from primary storage and/or from the electronic storage units containing documents that were previously migrated for non-disposition purposes. This model meets information management requirements since it allows for non-disposition migration but has been rejected for two reasons. This model would pose grave administration difficulties for the Information Manager. The Information Manager would eventually have a considerable quantity of electronic storage units to manage, maintain, locate and re-load when documents that have been migrated for non-disposition purposes must be moved back to primary storage for use. Also, documents that were migrated to electronic storage units for non-disposition purposes with a rated longevity of less than the retention period of the documents would have to be non-disposition migrated a second time to avoid loss of information thereby creating undesired technical work for the administrator.

The second reason for rejecting this model involved technical complexity in implementing this solution. The disposition migration process would have to physically move documents, with attached profiles, from off-line electronic storage units containing documents that were migrated for non-disposition purposes to electronic storage units designated for disposition migration resulting in a complex set of references to cover what documents were located where, and on which electronic storage unit. Also, it would be difficult to maintain the document content search capability for documents that have been migrated to electronic storage units for non-disposition purposes. In addition to these problems, it is doubtful that the complex back-referencing to the other portions of the CIMA required to implement this model would be possible to define successfully.

MODEL 3

In the third model, non-disposition migration is carried out to an on-line media rather than to off-line electronic storage units. This on-line media, referred to in this report as "secondary storage", should allow documents which would have been migrated to off-line electronic storage units in the two previous models, to reside on-line in "external storage". The only time documents would be written out to electronic storage units would be at the time of final disposition migration. Disposition migration implicitly assumes that the documents are being released from the organization's responsibility, and thus are of no further concern to the CIMA.

This third model is recommended for development for several reasons. First, the Information Manager would only need to be concerned with off-line electronic storage units at the time of final disposition migration, and not at the time of non-disposition migration. Second, the model would still allow for non-disposition migration, a definite requirement of the Information Manager. Also, the disposition migration process would be easier to define and administer, for the following reasons. Documents would not have to be moved from one electronic storage unit to another, thereby streamlining the process. It would also be easier for the CIMA to keep track of migrated documents because they would still be on-line, in secondary storage. The fourth reason for selecting this model involves the reduction of the time required by the Information Manager to carry out daily maintenance and management of electronic storage units. In addition, this model is more technologically feasible, because the document content search capability of documents migrated for non-disposition purposes would be much simpler to accomplish.

PREMISES

The selected EFM model, Model 3, is based on the following premises.

1) The cost of on-line storage is estimated to be 1 cent per page, and falling with time. Therefore, agencies should be able to afford to install on-line secondary storage that will meet their requirements.

2) The rate of filing is estimated to be 2 documents per day per user, at 5000 bytes of storage per page, and an average of 2 pages per document. Annual storage capacity required per 100 users can therefore be calculated as follows.

200 working days/year, 20KB/user/day = 200 X 20KB =

4MB/user/year = 400MB/year/100 users

3) It is acceptable for the Information Manager to use the rate of activity as the basis for non-disposition migration of documents to secondary storage.

4) It should be possible for any user, at any time, to search both primary and secondary storage for specified documents.

5) The CIMA will treat secondary storage as if it were primary storage, in that all daily CIMA operations should be able to be carried out on the documents in secondary storage with one significant exception, that is, **FILING**. Filing into secondary storage should not be allowed for two main reasons. First, secondary storage will no doubt be slower, and it would defeat its purpose if users were permitted to file into secondary storage. Secondary storage is intended for less active documents.

6) Approximately 5% of documents in most agencies are considered to be historical and therefore eligible for transfer to the National Archives.

A Diagram of the EFM Model has been included in Appendix D. The Information Manager should be able to specify limiting parameters that the CIMA will use in selecting documents for non-disposition migration to secondary storage. These are shown in the diagram as the box representing "Operator-Specified Parameters". Secondary storage is shown as a "black box".

In order to retain the generic quality of these functional requirements, secondary storage must be represented as independent of technology. The CIMA should simply send documents into and out of secondary storage with no assumptions as to the particular configuration of secondary storage. Different technological platforms within organizations will dictate the configuration of secondary storage in that agency. For example, in simple PC-based technology platforms, secondary storage might be merely an additional disk drive or dedicated server. On a mini or mainframe platform, secondary storage might be removable disk platters of fixed storage sub-systems, shown in the diagram as "Storage Unit 1 ..... n". Secondary storage could even be a proprietary storage sub-system, as is the case with imaging systems and optical jukeboxes. In all of these cases, a controlling sub-system, proprietary to the system vendor, would control the actual storage media. The CIMA would have to communicate with that storage subsystem via a vendor-specific interface module (VSIM).

The concept of secondary storage is crucial to the EFM model. In order to allow for any foreseeable technological implementation of modern storage systems, the following assumptions will be made about secondary storage.

1) The CIMA should be able to write documents and profiles directly to secondary storage, as if it were primary, on-line storage.

2) The CIMA should not be concerned with off-line electronic storage units when writing to secondary storage. If the secondary storage sub-system in fact handles removable media, such as optical platters, the proprietary controller should manage this internally.

3) The storage media *must be erasable*. The proper treatment and handling of security-classified documents is extremely complex and made even more problematic with a non-erasable media. Also, there are uncertainties surrounding the legal implications of using non-erasable media for classified or designated information. In addition to these two points, it is likely that erasable technology will become more prevalent in the next few years and will therefore not be a problem.

4) The secondary storage device or sub-system should independently manage the physical removal of CIMA-controlled documents to and from external storage units, if it uses removable media. The CIMA should not be expected to concern itself with the particulars of how the sub-system accomplishes this task.

5) Secondary storage will be assumed to be slower than primary storage, by a factor of ten.

6) The CIMA should be able to access all documents residing on secondary storage at all times.

7) The CIMA should never concern itself with available space on secondary storage. In other words, the CIMA should assume there is always sufficient space on which to migrate documents for non-disposition purposes. It should be the responsibility of the secondary storage sub-system to manage its space availability.

8) It should be possible to physically remove one or more electronic storage units from secondary storage, if in use, to enable the Information Manager to send it to an off-site facility for storage. The CIMA should not play a role in this extraction. It should be up to the secondary storage sub-system to accomplish this task.

9) The secondary storage sub-system should communicate the availability of each document to the CIMA. "Available" would mean that the CIMA could proceed to load the document immediately, without any time delay. If the document was unavailable, it would imply that operator intervention would be required to load an off-

line electronic storage unit.

10) Versions of the same document should be allowed to be split across primary and secondary storage.

The disposition migration process draws upon a set of rules, as set out in this report, and recommends a media format from a pool of media formats defined in the CIMA. During the disposition migration process, documents are drawn from both primary and/or secondary storage. Documents are segregated onto electronic storage units by retention period, and whether they are for destruction or transfer to the National Archives.

Therefore, EFM must define two substantive processes, non-disposition migration and disposition migration.

> *Non-Disposition Migration* works in both directions, i.e. OUT (from primary to secondary), and IN (from secondary to primary).
>
> *Disposition Migration* works only in the OUT direction, i.e. it is not reversible. This process cannot be influenced by the Information Manager except that the Information Manager can define the pool of media formats it draws upon.

The concept of efficiency is fundamental to this model. One of the reason for migrating documents to secondary storage is to relieve space congestion on primary storage. Primary storage is assumed to be more expensive, limited in size, and subject to congestion. It is likely that user response time will increase in some proportion to the quantity of documents stored on primary storage. Secondary storage is assumed to be less expensive, larger in capacity, and quite likely to be slower in retrieval, particularly in the case of removable or optical technology. Therefore, the Information Manager should use the EFM component as a tool to maintain the efficiency of primary storage by continually migrating as many documents to secondary storage as is practical.

In this model, the Information Manager should migrate documents to secondary storage based on a decrease in activity. The Information Manager's objective should be to maintain only documents that are still relatively active in primary storage.

As previously mentioned, the main elements of a document, as defined by the CIMA, include document content, the document profile, and document content search capability. All CIMAs, by definition, must have document content search capability. It is assumed that users will not tolerate a reduction in search capability for documents that have been migrated to secondary storage. Therefore, the document content search capability must be maintained for all documents that have been migrated to secondary storage. The CIMA users should be able to search for documents by content, by profile, or by file location. They must still be able to accomplish this for documents in secondary storage. However, it is assumed that, as explained earlier, there are technological compromises inherent in secondary storage. Searching for documents in secondary storage will likely be slower than on primary storage. The CIMA should therefore provide the following two types of searches.

> A *Regular Search* should only involve documents on primary storage. Information on secondary storage would not be included in the search. This type of search would allow a search of the profile, document content and file location.

> An *Extended Search* should include both primary and secondary storage. The extended search would include the search of profiles, document content and file location. This type of search would be a much less efficient search. As previously mentioned, a search on secondary storage is assumed to take up to ten times longer than a search on primary storage.

Given the present state of technology, maintaining the document content search capability of documents that have been migrated for non-disposition purposes may prove to be quite challenging for several reasons. First, the index overhead which must be maintained on primary storage for documents that have been migrated for non-disposition purposes is a prohibitive 20-50%. Second, in the event of an index failure, the index must be re-generated. In order to re-generate the index, all documents that have been migrated for non-disposition purposes must be read sequentially. This would be difficult and time-consuming if the documents themselves were stored on secondary storage.

Document **C**ontent **S**earch, or **DCS** is currently the best method to provide high quality retrieval capability in the midst of poorly specified subject lines, file numbers, and keyword fields. A superior replacement for DCS has been proposed, but is three to five years away from commercial feasibility. This new technique will be referred to as **TBS**, or **T**hesaurus-**B**ased **S**ummation. With TBS, documents are read by a linguistically intelligent process that draws upon a carefully-maintained agency-specific thesaurus, and approximately six words are generated per page of text. The words are theoretically representative of the true content of the document, and in conjunction with the thesaurus, deliver an extremely high level of retrievability without the requirement to store the overhead required by DCS. When TBS becomes commercially viable, it is recommended as the best alternative for document searchability.

16.6 NON-DISPOSITION MIGRATION, OUT

When the CIMA migrates electronic documents, it should do so in three stages. Each of the three stages is broadly defined below. Later, the various steps within each stage will be detailed.

STAGE 1 *BATCH DEFINITION* The process and rules of defining the batch of electronic documents for migration and determination of the type of migration. Once defined, the batch can be processed and the documents migrated.

STAGE 2 *BATCH PROCESSING* The processing that the CIMA carries out on the defined batch to prepare it for physical migration.

STAGE 3 *MIGRATION* Actual physical migration of the batch.

SPECIFYING NON-DISPOSITION MIGRATION CRITERIA

The operator should be obligated to specify criteria that the CIMA should use in conducting a non-disposition migration. The CIMA should then carry out the migration on those documents that have met the specified criteria. Each non-disposition migration process should begin with the selection of a *criteria set*. A criteria set should be comprised of a time frame with a start and end date, and one or more criteria from each of the following criteria groupings, based on the needs of the Information Manager at the time of migration.

INCLUSION CRITERIA Each item specifies documents that are to be included in the non-disposition migration. Only one criteria from this group is allowed per migration.

EXCLUSION CRITERIA Each item specifies documents that are to be excluded from the non-disposition migration. Any number of criteria from this group may be specified.

QUALIFIERS Each item specifies additional conditions that the documents must meet in order to be eligible for the non-disposition migration. Generally, these qualifiers are access-based. Only one qualifier is allowed per migration.

Each of these three groups of criteria is described below.

INCLUSION CRITERIA

1) All documents where the document age is greater than the active retention period.

2) All documents where the document age is greater than a specified period of time.

3) All documents where the expiry date is less than or equal to the present date. In other words, all documents where their expiry date has been met. This is the tool that would be used to remove expired documents from primary storage.

4) All versions of documents but not including the most current version.

5) Inclusion of specific files. In all cases, files should include all sub-files. File inclusion should be expressed as a single file number, a list of file numbers or a range of file numbers.

6) All documents which fall in a file with a specific component. The component may be any level of the filing system hierarchy, from primary to quinary and can include or exclude sub-files.

EXCLUSION CRITERIA

1) Specific files may be excluded. In all cases, files should include sub-files. File exclusion should be expressed as a single file number, a list of file numbers or a range of file numbers.

2) Exclusion by security level should also be possible. One or a combination of unclassified, protected, confidential, secret or top secret or other unique security classification levels, should be available for selection.

3) All sub-files under a specified file number should be allowed for exclusion.

4) Documents may also be excluded by a specified originator. The operator should be able to specify documents via the FROM field in the document profile.

5) All documents which fall in a file with a specific component. The component may be any level of the filing system hierarchy, from primary to quinary and can include or exclude sub-files.

QUALIFIERS

1) Less than a specified number of accesses per specified non-disposition migration period.

2) Less than a specified number of accesses after a certain date.

3) Less than a specified number of accesses prior to a certain date.

BATCH STRUCTURE AND HANDLING

Once the CIMA has selected the documents for non-disposition migration, a list of all selected documents, constituting a batch, should be written out to disk. The operator should be able to optionally modify the batch. Physical non-disposition migration should subsequently be a simple instruction to proceed, whereupon the CIMA should physically migrate the documents contained in the batch to secondary storage. A batch should be for a single migration type, that is, it should not be possible for a portion of a batch to be intended for non-disposition migration and another portion for disposition migration.

Once a batch has been created by the Information Manager using a combination of inclusion and exclusion criteria and qualifiers, the Information Manager will no doubt require some batch review capabilities to ensure that the batch is satisfactory prior to migration. The Information Manager may require the ability to identify documents contained in policy files, for example. Some Information Managers may wish to stop the CIMA from migrating policy documents contained in a batch, since their access rates can be unpredictable. Also, when examining a batch, the Information Manager should be able to see the documents in the context of their file titles. Document numbers are virtually meaningless to both the user and Information Manager and user-generated document subject lines are often ambiguous. Placing the documents in file context would help the Information Manager assess their validity for inclusion in the batch. It should also be possible to view the access rates of documents in the batch to determine whether or not a document is ready for non-disposition migration. The Information Manager should be able to select an option to view the entire history of a document, including all versions because what is NOT included in the batch may be relevant in deciding what documents to migrate to secondary storage. The Information Manager should also be able to view the TO and FROM fields for each document contained in the batch in order to decide whether to migrate or not.

In order to meet these requirements, the batch structure and review capabilities should have the following characteristics. There should be two views of a batch, a summary view which contains only the documents to be migrated along with their corresponding file titles for context purposes, and a detail view which would account for all documents in corresponding files with their associated versions, whether or not they have been selected for non-disposition migration. This detail view should also contain additional document profile information, and account for all versions of the documents in the batch.

Also, the operator should be able to direct the batch output to printer, screen, or disk in plain ASCII format. There should also be a capability within the summary view to display a full detail view of all documents in a specified file or file range within the batch. When the Information Manager views either the summary or detail modes on screen, the system should display the full file title parentage on the screen on demand. When printed to paper, the system should include file titles with full parentage for both modes. All documents contained in a batch should be displayed and/or printed with file titles, in file-number order. The file title and number should appear once for each group of documents in any given file.

PRIMARY STORAGE SPACE REDUCTION

The CIMA should supply the Information Manager with a means of freeing up an arbitrarily specified amount of primary storage space by using the non-disposition migration function, migrating documents to secondary storage. This particular criteria for non-disposition migration is mutually exclusive of the above criteria, since the CIMA is being asked to migrate a sufficient quantity of documents to free up the specified space, based on pre-defined internal algorithms.

The operator should specify the following three parameters for the CIMA to evaluate documents according to internal algorithms. Estimates, and not actual byte counts should be used because of the technical difficulties expected to occur with maintaining exact byte counts. The CIMA would have to constantly monitor byte counts, which could vary among documents of different structures. An estimate would serve the purpose nearly as well and would not impede system performance in any way. These estimates will likely be different from organization to organization and should therefore be configurable.

ACCESS SPAN A specified number of months representing a span of time. This is the period the CIMA will use to evaluate the least number of accesses in the specified time frame.

BYTES PER PAGE The operator should specify the average number of bytes per page for all documents in the filing system. The CIMA will use this figure to estimate the amount of space it will free up, by multiplying it by the estimated number of pages per document, and the number of documents selected for non-disposition migration.

PAGES PER DOCUMENT The operator should specify the average number of pages per document, for all documents in the filing system.

For every document selected for non-disposition migration, the CIMA should multiply the Pages Per Document by the Bytes Per Page to determine the estimated number of bytes to be freed up by migrating the document. The CIMA could therefore determine the total number of documents required for non-disposition migration in order to achieve the target amount of space reduction. The system should then proceed to run through each of the algorithms, selecting documents until the desired quantity has been reached. The system should then stop and declare the batch for non-disposition migration complete. The Information Manager must realize however, that the space freed by migration may not be an exact byte-for-byte match of the space requested due to the estimates used in the processing.

The algorithm used in this process is based on six levels of *qualification*. Each level contains a single qualifier that ranks all documents in descending order of qualification for non-disposition migration. The CIMA should evaluate all documents in order, from level one to level six. Within each level, different criteria should be used to determine when to stop selecting eligible documents, and proceed to the next level. The CIMA should pass through as many levels as required to select the required number of documents for non-disposition migration. If after passing through all six levels there are not enough documents to meet the desired space requirements, the CIMA should pass through all six levels again, this time with different criteria that should qualify a greater quantity of documents for migration. The CIMA should pass through the levels as many times as necessary to select the required quantity of documents for non-disposition migration. After the desired quantity has been reached, the non-disposition migration batch should be written out.

The six levels of qualification are:

1) File term has been met. That is, the document age is greater than the total of the active and dormant retention periods.

2) Oldest Versions. Each document's versions are ranked by age.

3) Closest to term, where the file has an approved schedule.

4) Closest to term, where the file does not have an approved schedule.

5) Oldest, by age of the document.

6) Least accessed during a specified date span.

7) Closest to the Primary Storage Expiration Period (PSEP) (see Back-referencing, below), where the PSEP is the period of time documents will be allowed to remain on primary storage after having been migrated to secondary storage.

BATCH MODIFICATION

Once the CIMA has defined a non-disposition migration batch, it should simply write the batch out to disk. The Information Manager may then manually alter the batch by adding new documents to it, deleting documents from the batch, or by modifying certain characteristics or attributes of documents within the batch. It is important to note that there should not be a need to modify the batch on an individual file basis in this case. If this needs to be done, the Information Manager should abandon the batch and begin again, based on different qualifiers, and inclusion and exclusion criteria. The three categories of allowable batch modification can be divided into three sections, namely additions, deletions and modifications.

It should be possible to add a new document number to the batch by specifying either all or a single version of a specific document number. It should be possible to delete a single version of a document or the full version series of a document. It should also be possible to delete all documents contained within a specific file number from the batch. Deletion of file blocks and single files and sub-files should also be permitted. The system should also allow the Information Manager to modify the batch by changing the status of the documents in the batch from Not Eligible for Migration to Eligible for Migration or vice versa. This should be available on an ad-hoc basis, whereby the operator would simply point to the document and reverse its status. A group of documents could have their status reversed according to the TO and FROM fields, when the document was last accessed, the security level of the document, the date the document was filed, or the total number of accesses on the document in question. The Information Manager should also have the ability to save or delete any particular batch.

Once the Information Manager is satisfied with the batch, s/he should simply instructs the CIMA to proceed with the migration, whereupon the CIMA should physically move the documents in the batch to secondary storage.

16.7 NON-DISPOSITION MIGRATION, IN

Non-disposition migration in the IN direction is functionally the same as in the forward OUT direction, with a few exceptions. The nature of such a migration implies that there is no need for a space reduction capability. Also, some of the non-disposition migration criteria are different than those described above. The Information Manager should be able to declare expiration periods for documents migrated back to primary storage. The CIMA would require a suite of capabilities for the Information Manager to manage these expiration periods. The availability of documents will also be a factor in that some documents may not be physically present on secondary storage. They may be stored off-line, depending on the secondary storage tool being used with the CIMA. In addition, the CIMA will likely require a suite of capabilities for the Information Manager to manage re-activation of files that have been migrated to secondary storage.

As the CIMA has been described above, users may retrieve directly from either primary or secondary storage. Therefore, whenever a user retrieves from secondary storage, the access count must be updated, as if it were in primary storage. This in turn means that the total number of accesses can increase for a given a document in secondary storage.

The stages of non-disposition migration IN are the same as those for non-disposition migration in the OUT direction.

PREMISES

1) The CIMA should record the fact that the migrated documents have been temporarily migrated back to primary storage. The CIMA should update all references to the file and documents, to reflect the presence of the documents on primary storage.

2) Once migrated back to primary storage, the original should be deleted from secondary storage. Technically speaking, a non-disposition migration in the IN direction should be a MOVE, not a COPY. Having two copies of the same documents in the system at once is deemed unnecessary and could cause confusion on the part of users.

3) It is estimated that documents migrated back to primary storage should be for short-term reference approximately 95% of the time. In these cases, the document should be considered to be in primary storage for a temporary period. Only about 5% of all documents migrated back to primary storage will be re-activated and maintained in primary storage.

INCLUSION CRITERIA

1) All documents where the document age is less than a specified time period.

2) All versions of a document up to but not including the most current version.

3) Inclusion of specific files. In all cases, files should include all sub-files. File inclusion should be expressed as a single file number, a list of file numbers or a range of file numbers.

4) All documents that were migrated to secondary storage on a specific date.

5) All documents migrated to secondary storage within a specific date range.

EXCLUSION CRITERIA

1) Specific files may be excluded. In all cases, files should include or exclude sub-files. File exclusion should be expressed as a single file number, a list of file numbers or a range of file numbers.

2) Exclusion by security level should be possible. One or more of unclassified, protected, confidential, secret or top secret or other unique security classification levels should be available.

3) All sub-files under a specified file number should be allowed for exclusion.

4) Documents may also be excluded by a specified originator. The operator should be able to specify documents via the FROM field in the document profile.

5) All documents which fall in a file with a specific component. The component may be any level of the filing system hierarchy, from primary to quinary and can include or exclude sub-files.

QUALIFIERS

1) Greater than a specified number of accesses per specified non-disposition migration period.

2) Greater than a specified number of accesses after a certain date.
3) Greater than a specified number of accesses prior to a certain date.

BATCH MODIFICATION

The batch modification process for non-disposition migration IN, back to primary storage, is similar to non-disposition migration to secondary storage. Once the CIMA has defined a non-disposition migration batch, it should simply write the batch out to disk. The Information Manager should then be able to manually alter the batch by adding new documents, deleting documents, or by modifying certain characteristics or attributes of documents within the batch. It is important to note that there should not be a need to modify the batch on an individual file basis. If this is the case, the Information Manager should abandon the batch and start another, based on different file qualifiers and inclusion and exclusion criteria. There are three categories of allowable batch modification processes.

As in non-disposition migration to secondary storage, it should be possible to add a new document number to the batch by specifying either all versions or a single version of a specific document number. It should be possible to delete a single version of a document or the full version series of a document. It should also be possible to delete all documents contained within a specific file number. Deletion of file blocks and single files and sub-files should also be permitted. The system should also allow the Information Manager to modify the batch by changing the status of the documents in the batch from Not Eligible for Migration to Eligible for Migration or vice versa. This should be available on an ad-hoc basis, whereby the Information Manager would simply point to the document and reverse its status. A group of documents could have their status reversed according to the TO and FROM fields, when the document was last accessed, the security level of the document, the date the document was filed, or the total number of accesses on the document in question.

One batch modification tool that will be required to migrate documents back to primary storage involves the primary storage expiration period which is described in more detail in the next section. The Information Manager should be able to change the Primary Storage Expiration Period (PSEP) from the default value, which would be applied automatically to all documents in the batch, and apply this change to a single document and all of its versions, all documents within a give file, all documents within a given file and all of its sub-files, all documents within a specified range of files, all documents within a list of single files. The Information Manager should also have the ability to save or delete any particular batch.

Once the Information Manager is satisfied with the batch, s/he should simply instructs the CIMA to proceed with the migration, whereupon the CIMA should physically move the documents in the batch back to primary storage on a temporary basis.

PHYSICAL MIGRATION

Back Referencing

Documents that have been migrated back to primary storage need not remain there. In fact, if pre-migrated documents were to remain on primary storage, the entire purpose of non-disposition migration would be defeated. The main reason for migrating documents in the first place is to minimize storage of documents in the primary space to ensure that search and retrieval speed remains at an optimum level. During each reverse migration, the Information Manager should therefore specify the number of days the documents are to remain in primary storage. This time period has been named the Primary Storage Expiration Period or PSEP. The CIMA should allow the Information Manager to configure a default expiration period, up to a maximum of 180 days, approximately 6 months, to be determined based on the unique requirements of each organization. Once declared, the Information Manager should not be able to delete the PSEP.

For each document migrated back to primary storage, several fields should be updated on the CIMA. The LOCATION field in the document's profile should change from Secondary to Primary. Also, the CIMA should calculate and display a field called EXPIRY DATE based on the PSEP for all affected document profiles. The CIMA should have a fully automatic, daily process that moves expired documents on primary storage back to secondary storage upon expiry of the PSEP.

File Re-Activation

All non-disposition migrations back to primary storage should be considered *temporary,* and should have a PSEP. However, on rare occasions, the Information Manager may be required to re-activate the file containing the migrated documents if the subject becomes active once again. In this case, the CIMA should have a new function called "RE-ACTIVATE FILE". The Information Manager should be able to specify a file number for re-activation, and the CIMA should change the file status from closed to open, and reset the PSEP of any documents contained in the file to 0, meaning they are no longer in primary storage on a temporary basis. The Information Manager should be able to re-activate a file on an ad-hoc basis while on primary storage, during non-disposition migration IN, that is, the Information Manager should be able to specify one or more files in the non-disposition migration IN batch to be re-activated. The Information Manager should be able to select files for re-activation by individual file number, list of individual files or range of files, files and all related sub-files, or specific components of a file or a range of files.

16.8 DISPOSITION PROCESSING

In the paper environment, the term *disposition* means to carry out the final action required on the records, as per an approved retention and disposition schedule. Each document is either be physically destroyed, or transferred to the National Archives for archival evaluation. In other words, the process of disposition, or disposal, consists of two sub-processes, or actions, namely *destruction*, and *transfer*. The CIMA should be designed in this same manner. This process or function should consist of two sub-processes called DESTRUCTION, and DISPOSITION MIGRATION. The migration referred to in this section is unique to disposition processing, and should not be confused with non-disposition migration. Disposition migration must have different rules and handling processes than would non-disposition migration.

Disposition processing in the CIMA should consist of three general steps, each which may be further broken down into sub-steps, as shown below. Each step is detailed later in the report.

DOCUMENT SELECTION All documents should qualify for inclusion in the disposition migration batch, according to two types of criteria. First, it should be mandatory that each document meet certain fixed criteria to qualify. Secondly, the operator should be given the option to specify certain operator-definable inclusion or exclusion criteria that would serve to broaden or narrow the list of qualifying documents to be included in the migration. The CIMA should select only those documents that have met both types of criteria.

The CIMA should then write out a "batch", an organized listing of eligible documents that met both the fixed and operator-definable qualifying criteria. The Information Manager should be able to modify the batch if required.

BATCH MODIFICATION In this step, the Information Manager should be able to exclude documents from the batch of documents selected for disposition processing. A number of exclusion criteria may be used.

DISPOSAL The disposal process should be broken into two major steps.

Migration Move the documents to be sent to the National Archives in the batch from the CIMA, either from primary or secondary storage, to off-line electronic storage units.

Destruction Delete the documents to be destroyed in the batch from the CIMA, either from primary or secondary storage.

The CIMA should execute these steps in the order shown. This means that during disposition processing, any documents that are to be migrated should be handled before any destruction is carried out.

PREMISES

1) Document profiles are considered equal in security to the documents they describe, and therefore must be handled with the same precautions.

2) According to the National Archives, the Information Manager should be transferring all historical documents that have met their retention period to the National Archives for archival evaluation. Should an problem arise that causes the planned shipment date to be missed, then a re-negotiation of the schedule is likely to take place.

3) Any documents designated as personal under the Privacy Act *must* be destroyed as per the authorized retention and disposition schedule, and should be destroyed on time.

4) Generally speaking, disposition processing should take place on a regular but infrequent basis. In a typical agency or large component thereof, disposition migration should be carried out once or twice per year.

5) Once electronic records have been migrated for disposition purposes, the originating agency should physically transfer those destined for the National Archives within a short period of time.

6) It is possible for either the originating agency, the National Archives, or an off-site storage facility to carry out the physical destruction of documents designated for destruction, assuming one of these three agencies has physical possession of the documents. However, an agency with an operational CIMA in place would have no requirement to ship electronic documents out for destruction since the process should be simple to complete within the agency. Therefore, the CIMA should not migrate documents destined for destruction to off-line electronic storage units. Instead, the CIMA should destroy these documents.

7) It is estimated that less than 20% of all agency records are sent to the National Archives for archival evaluation. This is expected to diminish to approximately 5% over the next several years.

8) At least 50% of all files have "event oriented" disposition dates, that is, the disposition date depends upon a certain condition. In general, there is a trend to schedule a greater proportion of records with these conditional retention periods.

9) The National Archives and the originating agency should generally have agreed in advance on a particular media type to use for the transfer of electronic documents.

10) Once the CIMA has migrated archival documents to electronic storage units, deleted those documents from the CIMA, and recorded the necessary details of the event, the responsibility of the CIMA ends.

11) Archivists insist that all media format types be accounted for within a date range of a given shipment. If information exists on paper, microfilm, and in electronic form, then all three media should be included in the transfer to the National Archives.

12) There is no requirement for an ad-hoc destruction process for two reasons. A PURGE function has already been described for special cases not related to the application of the schedule, and the selection criteria of the disposition process should be flexible enough to allow the Information Manager to narrow the selection down to a single file, thus this process should be able to be used an as ad-hoc tool for destruction if necessary.

DISPOSITION PROCESSING RULES

There must be a set of rules to govern affect disposition processing within the CIMA. These rules are detailed below.

1) All documents within each linked file must be migrated together in a disposition migration.

2) All media formats related to a specific topic known to the CIMA must be included in the migration.

3) The date spans of electronic document sequences must correlate to the date spans of non-electronic and external documents.

4) A unique agency identifier must be recorded as part of each migrated document's profile.

5) All versions of a given document, from the oldest to the most current, must be migrated together. The versions of a document are known as a *document series*.

6) An authority number must be present before a document can qualify for disposition migration.

7) The disposition migration process must not be able to transfer documents to be destroyed to off-line electronic storage units.

8) Once a document series has been disposition processed, the document number should never be re-used within the CIMA. This precaution is needed to avoid potential duplication problems at the National Archives and with other CIMA users.

9) If the documents fall within a linked file and its status is closed, all documents in the file must be processed together.

DOCUMENT SELECTION

All documents must meet certain fixed criteria to qualify for inclusion in the disposition batch. In order to qualify, a document must have met its disposition date. In other words, its age must be at least that of the disposition date at the time the document is evaluated for disposition. The disposition date is the date on which a document has met any required conditions to qualify for disposal under the terms of an authorized retention and disposition schedule.

In order to evaluate a document for eligibility in the disposition batch, the CIMA must know the disposition date of its host file. For clarification purposes, all files can be broadly categorized into two categories according to the method of determining their disposition date.

Chronological Documents in this type of file qualify for disposition after their age exceeds a pre-determined, fixed period of time. Generally, a document qualifies after its age exceeds the file term, that is, the sum of its active and dormant retention periods.

The disposition action for these files should be either destroy, based on a chronological date, or historical, transfer to the National Archives based on a chronological date.

Conditional This type of schedule involves some "condition" used to determine when a document qualifies for disposition. For example, documents may qualify "five years after contract closure", or "two years after sale of equipment". In order for the CIMA to know when the documents qualify for disposition, the Information Manager must manually record the disposition date as part of the file's attributes when the date becomes known. The CIMA would then be able to evaluate the documents in the file for qualification by comparing their age with the disposition date.

The disposition action for these files should be either destroy, based on a certain condition, or historical, transfer to the National Archives based on a certain condition.

As mentioned earlier in this report, not all files will have a known disposition date. The CIMA should not qualify any documents for disposition migration without a known or a calculable disposition date.

In order to qualify for inclusion in the disposition batch, each document series must meet the following criteria.

1) The host file must have an authority number recorded.

2) If the host file is a linked file, its STATUS must be closed.

3) If the file has a CONDITIONAL disposition date, its STATUS must be closed.

4) The age of the document series must meet the disposition date, according to one of the following algorithms.

   If the document is in a chronological file, the age of the document series must be greater than or equal to the present date minus the date the most recent version was filed. If the document is in a conditional file, the age of the document series must be greater than or equal to the present date minus the disposition date.

All documents that meet the above fixed criteria are theoretically qualified for inclusion in the disposition batch. However, the Information Manager may further specify additional operator-definable criteria that documents must meet in order to qualify. The Information Manager should be able to specify any of the following criteria as either inclusion or exclusion criteria, and should be able to specify any combination of inclusion and exclusion criteria. The CIMA should allow the inclusion or exclusion of documents that fall within a particular date range, a specific file, file range or list of files, documents that fall into a file with a Personal Information Bank number, documents within a specific file and all related sub-files, or documents that are contained in a specific file component across the entire filing system hierarchy.

The CIMA should evaluate all documents to determine if they meet the criteria. Those that do meet the criteria should be included in a disposition batch, which the Information Manager should be able to view and alter if required.

BATCH MODIFICATION

Once the disposition batch of qualifying documents has been written out, presumably to disk, by the CIMA, the Information Manager should be provided with a convenient means of reviewing the batch and excluding any documents. This would be a tool of convenience for the Information Manager since it should be faster and easier to exclude exceptions from an existing batch than to re-generate a replacement batch beginning with an entirely new set of criteria. The CIMA should present the Information with a view of the batch by file number, document number, version number and the date each was last accessed.

The Information Manager should be able to modify the batch, within the rules of disposition processing, as often as required until a satisfactory batch has been defined. The Information Manager should be able to use the same criteria as for the initial selection, but the criteria should used only be allowed to exclude documents, not to include them. The CIMA should take no action on a batch until specifically directed to proceed with the disposition processing of the batch by the Information Manager.

DISPOSAL

Physical disposal takes place on the documents as defined in the disposition batch. The CIMA should first migrate all documents to be transferred to the National Archives to off-line electronic storage units, then proceed to destroy all documents in the batch designated for destruction.

DISPOSITION MIGRATION

CIMA should go through the following steps to migrate documents for disposition purposes.

1) The CIMA should prompt the Information Manager for the most appropriate media format which must correspond to one of the media format types known to CIMA from the media format pool.

2) The CIMA should calculate the quantity of electronic storage units required for the given media format, and inform the Information Manager of the number.

3) A series of labels, each sequentially numbered, should be printed for application to each electronic storage unit. Each label should show the unique electronic storage number, the highest security level for documents on the electronic storage unit, the originating agency name, the number of bytes occupied, the date span of the documents, the migration date, and a shipment identifier.

   The electronic storage unit numbering scheme should include the year the documents on the electronic storage unit were migrated followed by a unique identifying number, beginning at 0001.

   Electronic storage units might also be segregated by security. For example, electronic storage units #1 and #2 may contain Secret documents, electronic storage unit #13 to #32 may contain Top Secret documents. The CIMA should also allow segregation of unclassified and protected documents. This would be necessary for electronic documents produced by secure systems, that is, TEMPEST-certified workstations and networks. The CIMA should be able to cluster a number of documents together, that contain one or more secure documents.

   At the present time, current technology does not allow the CIMA to safely destroy information designated or classified under the Security Policy of the Government of Canada. The policy calls for the destruction of the recording medium itself. Primary storage, and potentially also secondary storage, in the CIMA simply cannot be destroyed. Storage in the CIMA is assumed to be conventional, sealed, internal magnetic media that cannot be readily removed and replaced. As a compromise, each time the CIMA writes out a electronic storage unit label, it should stamp the security level of the highest document on the electronic storage unit's label.

4) On the Information Manager's command, the CIMA should proceed to move all documents for migration to National Archives out to

electronic storage units. It should be noted once again that the documents are to be copied to the electronic storage unit, then erased from the filing system, that is moved, not copied. Also the profile of each document should be written to the electronic storage unit in plain flat ASCII format. There must be a means of logically associating each document with its profile. Each non-electronic document should have its SUMMARY field recorded as a document, as if it were an electronic document. Finally, the originating agency identifier should be stored with each document's profile. This is for the benefit of the National Archives who must be able to distinguish between documents from various originating agencies.

Once all electronic storage units have been written out, the CIMA should proceed immediately with destruction of the documents authorized to be destroyed.

DESTRUCTION

In this step, the CIMA simply erases each document in the batch which has been authorized for destruction from the host filing system, either in primary or secondary storage.

POST-DISPOSITION PROCESSING

After a batch has been fully processed, the CIMA should keep a record of the batch indefinitely as a detailed audit trail of the disposition. This can be referred to as the *disposition record*. At any time, the Information Manager should be able to generate a report of the disposition. The Information Manager should be able to delete the detailed record of any specified disposition at any time, add an accession number to any specified disposition record at any time, or query all the existing disposition records to extract any desired information. The Information Manager should be able to specify a query consisting of either the disposition processing date, or the file number, to determine which disposition contained a certain file number.

If, following disposition processing, a file is left completely empty of documents, the CIMA should record the following information as an audit trail; file number, date the file was emptied, and the disposition action, and delete the file from the CIMA, assuming it has no sub-files.

ORGANIZATION OF MIXED-MEDIA SHIPMENTS

The hierarchical breakdown of a shipment of electronic records should be as follows.

SHIPMENT
CONTAINERS
ELECTRONIC STORAGE UNITS
FILES
DOCUMENT SEQUENCES
DOCUMENTS
VERSIONS

When packaging the documents demarcated by a document sequence, each group of documents of a given media format should be packaged in an independent series, that is, all boxes of paper should begin with container number one. Documents of a different media format should have another container numbering series, again beginning with one.

16.9 SUPPORT FUNCTIONS

SEARCH CAPABILITY

A key objective of the CIMA is to provide the user with the ability to find a document with the same ease and performance whether or not it has been migrated. It should be noted that finding a document does not involve the same processes as retrieving a document. The CIMA may be able to quickly locate a document and provide the user with many other details about the document, but, depending on the secondary storage facility in use, the document itself may physically reside on an electronic storage unit in another building, which may take days to retrieve.

The search capability provided by the CIMA should be available for all documents, both those documents on primary storage and those that have been migrated. Thus, in searching documents, the processes of the CIMA should be transparent to the user until the user wanted to actually view or retrieve the document. If the document had been migrated, the user may be unable to immediately view or retrieve it.

All documents in a SEARCH result list must therefore contain a visual indicator to show that a Primary Storage Expiration Period exists for that document. Also, for those documents with a Primary Storage Expiration Period, the calendar date of expiry, as calculated by the CIMA, should be shown.

PREMISES

1) It is possible for a version of a document within a series of versions to be missing, either because it was purged by the Information Manager or destroyed as per an approved schedule.

2) All document profile fields, and document content search capability, should be available for all documents on both primary and secondary storage.

3) The user should be able to view the profiles of all migrated documents, in lieu of the actual document content.

4) The user should be able to request that a migrated document be retrieved.

5) Approximately 5 to 10% of all work is expected to be long-term, where the user would require that all versions of documents be fully accounted for.

VERSION ACCOUNTABILITY

Versions of a document may exist in either primary or secondary storage, and versions may be missing due to a PURGE or DESTROY operation. Therefore, the CIMA should have a means of accounting for all versions at any particular time. As previously mentioned, a regular search should include only documents held in primary storage space, and there need not be accountability for "missing" versions, that is, those not on primary storage. There should be a second search capability however, called EXTENDED SEARCH, that should provide additional information to the user including those documents located in secondary storage space, and all missing versions of documents.

VIEWING DOCUMENTS

There are two types of processes that users would require for electronic documents.

View/Print This function should display the document on the screen or send it to the printer. This process must be fast, and should not require the use of the native application. It should be non-invasive in that this process should not alter document content.

Viewing would generally be accomplished through a software process called a "viewer", which quickly displays the document on the screen, in a close representation of how it would look if it were loaded into the original application software used to create it.

Edit If a document is stored in native format, the original application software must be loaded to edit the document contents. The application version must be precisely matched to the particular native format.

Document viewers are useful in the case where no universal document format exists. Viewers make it fast and easy to display the documents in their native formats. However, they should not be seen as a substitute for the universal document format because the benefits of the universal document format outlined previously are not offered by viewers. The only reliable, meaningful long-term solution to the real-life multiplicity of document formats is a universal document format. Ideally, all government agencies would adopt a single format, and strive to maintain it, modernize it, and build and encourage the development of support tools for it.

16.10 MEDIA CONSIDERATIONS

When the Information Manager migrates a series of electronic documents, s/he should be required to select a particular media format from a selection of available formats. The CIMA should not enforce the selection of media formats in any way. Instead, only the following recommendations should be made.

1) The Information Manager should choose a media format that has a longevity that matches or exceeds the retention period of the documents to be stored on it. This should reduce the need to replace the media prior to the end of the retention period.

2) Once a document has been migrated for disposition purposes, the copy on the electronic storage unit becomes the only copy in existence, unless the CIMA filing system is backed-up on a regular basis and the back-ups safely maintained for a period of time. Therefore, the CIMA should prompt the Information Manager for either one or two copies of each electronic storage unit at disposition migration time. The duplicates of the migrated documents would serve as protection should anything happen to the electronic storage units while in transit to the National Archives.

3) Each electronic storage unit should contain the document profiles along with the documents themselves. All profiles on a given electronic storage unit should be stored as a single, contiguous machine-readable plain ASCII file.

The CIMA should propose a scheme for describing media formats. If all agencies used this method, all parties would be able to reliably communicate about their media, and be fully aware of their particular media characteristics.

The following media model simply lists all of the elemental characteristics of electronic media, and defines the ranges into which these characteristics fall. To describe a particular brand/make of media, one would simply describe it by identifying the items on the list that it satisfies.

Each actual media type therefore would be described by use of this standard model and would be assigned an agreed-upon media "code" defined by the CIMA as a media format. The adoption of such a model would help standardize the media formats in use across government.

In the list of characteristics below, each selection within a characteristic would be represented by an agreed-upon code.

The following are the characteristics of the media model.

1) Physical form. Roughly describes the physical characteristics of the media. Examples:

5 inch floppy
3 inch floppy
5 inch optical
12 inch optical
3/4 inch magnetic tape

2) Storage Format. Method by which data is stored on the particular media. Examples:

High Sierra (5" optical storage of 1990)
MS DOS 4.x
MS DOS 5.x
Apple MAC 2.x and 3.x

3) Accessibility. One of:

Online -- no operator mount required
Offline -- operator mount required
Online -- robotic mount required (jukebox)

4) Longevity. Defined as UNATTENDED shelf life, when stored at recommended manufacturer conditions. One of:

< 2 years
2 - 5 years
5-10 years
10-20 years
20-30 years
30-100 years
100+ years

5) Capacity. Defined as characters per electronic storage unit. One of:

1 - 1,000,000
1,000,000 - 10,000,000
10,000,000 - 100,000,000
100,000,000 - 500,000,000
500,000,000 - 1,000,000,000
1,000,000,000 - 5,000,000,000
5,000,000,000 +

6) Access speed. Expressed as seek time, or the average time it takes to locate a specified document. Valid selections are:

< 1 millisecond (ms)
> 1 ms, < 10 ms
> 10 ms, < 100 ms
> 100 ms, < 1000 ms
> 1000 ms

It would no doubt be helpful to provide an example of the use of such a model. Assume that common 5 inch MS-DOS PC floppies are assigned media format #33. This media format would be defined as follows.

Physical form 5.25 inch
Storage format MS-DOS 5.0
Accessibility - Offline
Longevity < 2 years
Capacity 1-1,000,000 characters
Access Speed > 1000 ms

17.0 AUDIT TRAILS

An audit trail is a log of events, maintained by the system, which records activity on certain pre-determined functions. The functions chosen should depend somewhat on the organization but some should be standard.

Access History

This would involve keeping an audit trail of access to each file within the system for retention and disposition management purposes. Access frequency is important in determining the value of the information and when the information should be migrated. Fields for this audit trail would include the date of access, the user name, file number, volume number, document number and the type of retrieval action taken, that is, view or actual retrieval.

Deletion History

Whenever a file, document or version of a document is purged or destroyed, the system should maintain an audit trail of the deletion and record the login name of the user executing the deletion, the date the information was deleted, and provide a field for comments where the user could enter meaningful information. It should be noted that the purge and destroy functions should be restricted to the information manager to be used for disposition and error correction purposes only.

Creation, Deletion and Transfer

The system should be able to keep track of the number of files created, deleted or transferred by date. These statistics would prove to be extremely useful for system performance analysis as well as for standard statistical purposes. Whenever a file, document or version is created or deleted, the CIMA should record the operator login name, appropriate comments, and the date of the action.

Unauthorized Access Attempts

The application should keep an audit trail of all unauthorized access attempts for security purposes.

Hardware and Software Errors

The system should track and log all problems, failures and errors relating to both hardware and software to facilitate application management.

Schedule Maintenance

The audit trail for schedule maintenance should record the date of change, the original and new schedule data, and the user login name for each change.

File Re-Activation

For each re-activated file, the CIMA should record an audit trail including the file number and the date of the re-activation.

18.0 REPORTS

In traditional automated systems, one of the most dynamic parts of the system design is the reporting function. Users constantly request new reports, changes to existing reports and amalgamation of different reports. Since report requirements vary greatly, these functional requirements outline only basic reports. A system feature should be the generation of report specifications that could be used only once or saved for later use. These reports could be configured by the application manager or the information manager. Users should also be able to utilize this feature for the generation of simple reports.

Certain pre-defined reports will no doubt be necessary. Records office operations reports have been in existence for a number of years and have evolved with user input. Many of these types of reports could be made available. The CIMA should provide a variety of reports, for administrative purposes. The CIMA should allow the Information Manager to send all reports to screen, printer or disk. If written out to disk, the reports should in flat ASCII format. The CIMA should also provide an ad-hoc reporting facility, whereby the Information Manager may generate a report of any fields, in any order, based on any data items recorded in the CIMA.

Some of these reports are outlined below.

AUDIT TRAIL REPORTS

Several reports would be required for audit trails. For example, a file access history report organized by user, date, file number, and document number would help the information manager with retention and disposition issues. A report of unauthorized access incidents would be useful for security reasons (see Section 17).

RECORDS CENTRE MANAGEMENT REPORTS

Several reports for the Records Centre Management portion of the system are required. They should be pre-defined reports that are required to support records centre operations. Upon entry to each report, the information manager should be presented with a form that would allow configuration of certain parameters such as start and end ranges, headers, column selection, etc. The report contents should be pre-determined and fixed where appropriate. Ad hoc reports should also be made available.

Some of the most common reports are outlined below.

Recall Summary

The recall summary is a list of all overdue files. In this report, the recall date must be greater than the present date. This report should include prefix, file number, volume number, charge out date, recall date and name of the user to whom the file was charged.

Recall Notice

This report involves generating recall notices to be sent to users for all overdue files. These notices could be sent electronically. If printed, the recall notice should be printed one per page since the notices must be sent to different users. Where the same user has several overdue files, the system should be able to print these notices sequentially, more than one on a single sheet, to avoid paper waste. The recall notice should be divided into two sections. The top section should bear the user's name and location, the present date, the file and volume number in question, the date the file was charged out, the subject of the file, the records office location where the file should be returned, and the records office contact name and telephone number. The bottom portion of the report should provide options for action including file attached; BF date; file was returned on; file was passed to, on; user signature block, telephone and date block to be filled in by the user.

BF (Bring Forward) List

The BF list report should include the BF date, prefix, file and volume numbers and name of the user requesting the file. The list should be sorted by BF date, then by prefix, file and volume number. The list should only contain BFs that meet a specified date, usually the present date or the next day's date. This should be predetermined by the needs and practices of the organization.

ESSENTIAL RECORDS REPORTS

A summary report listing all records flagged as essential by category would be extremely useful in the management of an essential records program.

DISPOSITION MANAGEMENT REPORTS

Shipment Inventory List

Shows all files being packaged to form part of a shipment. Sorted by container number, file number, volume number, date spans, media format, file title.

History Report

For inclusion in a shipment of information to the National Archives. Shows the entire history of a file, from its creation to deletion. Should include file number, previous file numbers and titles and remarks.

Document Sequence

For a given file sequence (or, multiple file sequences in a batch), this report should identify the various media formats included in a given file sequence. Would be used prior to actual scheduling.

Records Quantity

This report would be used in forecasting resources necessary to prepare a shipment, or to plan a shipment to match available resources. Should be by document sequence within a specified date span, then by media format, and number of volumes, documents, etc. per media format.

Storage Summary

This report should estimate the amount of physical storage space that would be freed up for a given migration operation. Paper space is estimated based on the presumption that one volume equals one inch of paper.

Inventory Listing

The inventory listing should be an ad-hoc report that the Information Manager would produce after identifying a series of documents for migration to the National Archives. It should be noted that all gaps in this report should be accounted for, via explanation codes outlined in this report. Gaps are considered gaps in either date span, or volume sequencing.

Holdings Stability

This report should list all migrated documents, and show the retention period as compared to the media format longevity.

Disposition Report

One report should be produced per shipment, includes file number, document and version numbers, file title, media format, disposition action, and disposition date. Information Manager should be able to add accession number at later date.

19.0 ARCHIVAL CONSIDERATIONS

Throughout the various phases of the IMOSA project and its predecessors, one of the fundamental objectives has been the preservation of the historical record. There should be the ability to manage all aspects of the lifecycle of information prior to its final disposition within a CIMA. If there is to be a complete management of the information lifecycle, there should be an archival component to a CIMA, either internal or external to the organization. This section of the report is intended to outline some of the considerations involved if records created in a CIMA environment are to be subsequently available and used in an archival context. A model of such a system with an explanation is provided below. The intent of this model is not to detail each step nor dictate what system should be adopted. Rather, it is intended to describe what tools an archivist would require to carry out the traditional archival functions on the information.

An A-CIMA, or Archival Corporate Information Management Application, as described in the model below, is a system which would have to be created in order to permit the National Archives to acquire and manipulate documents created and managed by federal institutions employing a CIMA. Such a system should possess several of the characteristics of the CIMAs used in those institutions, as well as include the procedures required by the National Archives for internal use. In short, this system would be intended to assist the National Archives in the acquisition, description, conservation, and research consultation of documents acquired from a organization using a CIMA. It should be emphasized that outside the National Archives, provincial or other corporate archives would also be interested in such an application when their clients turn to a CIMA for the management of their own information.

TRANSFER CONSIDERATIONS

Government agencies should transfer documents which have been managed by a CIMA to the National Archives on a regular basis according to retention and disposition schedules. The electronic storage units, as described in this report, are the final product of the CIMA, the medium on which information is transferred to the National Archives. The document profiles, the index information from the search tool, the electronic documents themselves, as well as the profiles for those documents found in other media, should be stored on the electronic storage units. At the time of acquisition it should also be necessary for the National Archives to obtain documentation about the system, particularly in those cases where the information transferred includes data files, and where the record structure and data dictionary are required for archival use.

It is also important to note that the issue of a universal storage format is central to the long-term use of electronic records. Conversion to such a format is necessary for an Archives to be able to preserve and use such information. With the adoption of a standard by the Government of Canada, this conversion would be able to take place prior to the transfer of the information to the National Archives. However, until such a time, the National Archives should convert records to such a format or negotiate a format to be used in transferring documents for archival retention in the disposition process with each institution.

In addition, issues surrounding media such as physical form and storage format, whether in DOS, Unix, or other, for example, should have to be addressed sometime prior to the actual transfer.

ACQUISITION PROCEDURE

Ideally, the National Archives should be in a position to recopy electronic storage units to a format which should meet long-term archival storage, via the use of an archival electronic storage unit at the time of acquisition. These archival electronic storage units would then be preserved in an Electronic Records Repository, as is the case at present, while the relevant non-electronic documents in other media formats such as paper, microfilm, cartographic records, and photographs, would be preserved separately under the proper conditions.

The primary attribute of the A-CIMA should be its capacity to manage the profiles of the electronic documents and the non-electronic documents transferred by the originating institution. In theory, these profiles would be copied from the archival electronic storage units at the time of acquisition, creating a database of all the records transferred to the Archives that had been managed by a CIMA.

This database would also permit control of the relevant documentation provided by the originating institutions when required. It should also be capable of transmitting relevant information to applications used as finding aids for use inside or outside of the National Archives. In such cases it would be important that the emerging "Rules of Archival Description" be respected so that these records are widely available in the Canadian Archival Community.

RELATED PROCEDURES FOR THE MANAGEMENT OF CIMA RECORDS

Another important characteristic of the A-CIMA should be the handling of the electronic documents created by a CIMA within a federal institution. Such documents consist of three inseparable parts, the document profile, the index file from the search tool, and the documents themselves. When a record is identified by its profile, it should be possible for the system to point to the archival electronic storage unit on which the document is stored. The archivist should also be able to adequately manage information on archival electronic storage units according to their different functions, such as the selection and destruction of files without archival value, responding to formal access requests and public use of the records, managing the archival electronic storage units within the life expectancy of the media format, that is, periodic rewinding or recopying of the media, as required. Thus, such a system would necessitate many operations requiring the movement of information to and from the archival electronic storage units. This movement of information is referred to as Archival Migration.

20.0 CONCLUSION

The functional requirements outlined in this report should serve as a base for the development of automated solutions for the management of electronic information in office systems. Issues relating to security, access, corporate memory, retention and disposition management, system configuration, audit trails and reports have all been addressed. As more organizations attempt to address these issues through the acquisition or development of software solutions, these functional requirements will no doubt be superseded, or at the very least, modified to include factors learned through more direct experience.

ANNEX 1

SELECTIVE GLOSSARY

SELECTIVE GLOSSARY

Access A measurement of activity of a an electronic document. An access is considered to have taken place when one of the following is performed on the document:

Viewed
Retrieved

A-CIMA Archival Corporate Information Management Application.

Active Retention Period The length of time information should remain active.

Age, Document The age of a document at any point in time is calculated to be [Current date - date filed].

Agency Identifier An alphabetic code used to denote the agency the documents originated from. Sample identifiers, DND, EMR, ENV.

Back-referencing Updating the references to migrated documents, either during or following either non-disposition migration or disposition migration.

Batch A batch of electronic documents, designated for migration. Consists of a specific format of file numbers, and document sequences. Electronic documents have to be batched before they can be migrated by the CIMA.

BF Entry A BF entry is a request by a user for a paper or electronic document to be brought forward in the future. Each entry records user name, BF date, file number, volume number, document number, document type.

BF Reminder An on-screen notification seen by the originating user at login time. It is a message indicating that a CIMA document has been brought forward for that day.

Block, File An inclusive range of primary file numbers, including all children files.

Charge-out Entry The series of data required for the processing of a charge-out such as user name, date, file number, recall date, volume number, etc. Records the charge-out event.

CIMA — Corporate Information Management Application. Refers to an automated system that manages electronic documents.

Component, File

A file component is one of the constituents of a complete file number. For example, the file 1100-23-2 has 1100 has the primary component, 23 as the secondary component, and 2 as the tertiary component. Used in specifying migration criteria.

Corporate Domain

A document storage space controlled and administered by the CIMA. Electronic equivalent of an institution's paper-based filing system.

Correlation, Date

A circumstance where the date span of a series of non-electronic documents not managed by the CIMA matches the date span of a document sequence. Required when preparing a batch of documents for migration.

DCS

Document Content Search capability. Does not constitute actual data, rather the capability of searching for any word/phrase within the body of the document. Typically implemented in most modern CIMAs via full-text retrieval technology. Generally provided by maintaining a so-called "content index" on primary storage, occupying 20-50% of the total aggregate document size.

Demarcation

The manner in which a portion of electronic documents within a given file are "set aside" as a contiguous series, for the following purposes:

> Matching the series with the volume start and end dates of the paper documents of the same file number.
>
> Arbitrarily designating a series for migration, irrespective of the paper documents within that file.

Destroy

To delete a document from the filing system, as per the proper procedures and with the authority of an approved schedule. Bears no relation to PURGE, where a document or version may be deleted without regard to the host file's schedule status.

Destroy Documents

Documents slated for destruction under the schedule.

Dispose/Disposal

Action of passing the document from the ownership of the originating agency to the ownership of the National Archives. Not to be confused with MIGRATE, which means little more than the physical transportation of documents from one physical location to another. Documents may be "migrated" to points within the originating agency, to off-site storage facility, or to the National Archives, but can only be disposed of to the National Archives. Not related to destruction.

Disposition Action

What should happen to a document, according to an approved schedule, after its retention period has elapsed. Represented by a code such as DET, referring to a specific action to take. The following standard Disposition Actions are common to most agencies.

> Destroy, chronological. Destroy when file term expires, no other conditions apply.
>
> Destroy, conditional. Destroy when a specified condition is met.
>
> Historical, chronological. Send to National Archives when term is met.
>
> Historical, conditional. Send to National Archives when a specified condition is met.
>
> Unscheduled. CIMA defaults to a five year Active Retention Period and a ten year Dormant Retention Period.

Disposition Date

Date that a file qualifies for disposition processing. Part of file profile. Used where a condition is used to determine when a file becomes eligible for disposition.

Disposition Migration

Migration of documents at the final disposition point, i.e. when the retention period has passed. Opposite of non-disposition migration.

Document Format

The manner in which a document's content, format information, and text attributes are organized for storage. Either a **native** format, such as WP, MS WORD, etc., or a **UDF** (Universal Document Format) such as ODA/ODIF.

Document Profile

Collection of data fields associated with a document.

Document Sequence A complete series of documents of a file, represented with a start date and an end date. Can be considered to be a "slice" of a file's documents, where the slice begins at a start date, and ends at an end date. For example, all documents of file 1200-2-2 from Jan. 1 1960 to Jan. 1 1970 would constitute a document sequence. In the case of a disposition migration, no document forms must be omitted from the sequence without explanations.

Dormant Retention Period The length of time information must be retained in a semi-active or dormant state prior to final disposition by destruction or transfer to the National Archives of Canada.

EFM Electronic File Migration. A set of features of a CIMA, specifying the migration of electronic documents.

Electronic Documents Documents that are entirely in electronic form, and are stored within, and managed by, the CIMA. A "profile" of information about the document, such as subject, author, etc. is associated with it.

Electronic Storage Unit Single physical unit of a particular record form, or a particular media. Cannot be further subdivided. For example, three floppy disks represents three electronic storage units of that particular record format.

ESU See ELECTRONIC STORAGE UNIT.

Expiry date Date a document is to be returned to secondary storage from primary storage. Derived by adding the Primary Storage Expiration Period to the date the document was migrated for non-disposition purposes from secondary storage.

External Documents Documents entirely outside the CIMA. They are not recorded within CIMA, nor does CIMA record anything about them. They may be of any form, i.e. electronic, paper, microfilm, etc. For the purposes of migration, particularly disposal to National Archives, external documents should be associated with, or bundled with, the electronic and non-electronic documents managed by the CIMA.

File Profile Collection of data fields associated with a file.

File Status File status is either Closed, or Open. If closed, a file is eligible for disposition migration.

File Term — Sum of the Active Retention Period plus the Dormant Retention Period.

IN

Direction of document migration from a less active to a more active state, from the originating agency's point of view.

Keep Documents

Documents deemed to have archival value, and should be migrated to the National Archives upon expiration of the retention period.

Linked File

Case file. File where all volumes must be kept together throughout its lifetime (i.e. until closed).

Look-Up

The look-up function searches the filing system for a particular file number or subject. The look-up function is used to find a file number based on a query of text in the subject title, or to position the user in a particular range of file numbers and their titles. It is used primarily to assist the user in assigning file numbers to documents.

Media Format

Refers to a unique physical form of a document. i.e. 5-inch IBM-format floppy, 9-inch reel tape, 12-inch optical platter.

Migrate

To physically move documents IN or OUT between active and less-active states, from one physical location to another. There are two types, non-disposition migration and disposition migration, as each type of migration has different requirements.

Migrated

Not stored on primary storage. Has been migrated by the CIMA. Only information *about* the document is stored on the CIMA and available for searching. In order for a user to retrieve the document, it should have to be migrated back to primary storage. Generally not available immediately. Opposite of *present.*

Migration Type

Either disposition migration, to the National Archives, or non-disposition, migration internal to the originating agency.

MGIH

Treasury Board policy entitled Management of Government Information Holdings.

Native Format

Method of organizing the content and format information of a document that is unique to the process used to create the document. The document creation process is required to edit/view the document. E.g. WordPerfect 5.1, MS Word 3.3, etc. Contrasts with UDF (Universal Document Format).

Non-disposition Migration

Migration of documents other than at the point of final disposition. Opposite of disposition migration. Often this means migration to the off-site storage facility, or perhaps within the agency itself.

Non-electronic documents

These documents are recorded by the CIMA, but are not stored within it. They are treated by the CIMA exactly the same as electronic documents. A "profile" of information about the document, such as subject, author, etc. is associated with it. An example might be a paper report, or a VHS video cassette.

Off-site Storage Facility

A location for records storage, on premises outside those of the originating agency. Usually a commercial records centre, or the Federal Records Centres run by the National Archives.

OUT

Direction of document movement from a more active to a less active state from the originating Agency's point of view.

Personal Domain

User's workspace. Any and all disk storage space available to the user, i.e. all addressable disk drives. This includes local floppy and internal hard disk storage, and any disk space available on a shared file server.

Pool, ESU

All allocated electronic storage unit numbers. The CIMA maintains a list of all used electronic storage unit numbers, and a record of what documents are stored on each. Each electronic storage unit number is unique.

Pool, Media Format

All available media formats that the Information Manager has defined for the CIMA. At migration time, the CIMA draws upon this pool of media formats to recommend which to use for document storage.

Prefix

Top level in a filing system. Composed of sections, primary subjects, files and sub-files. Consists of a name and a title represented by an alphabetical or numerical code.

Present

Present on primary storage, immediately available in its entirety. No need to migrate from secondary storage to retrieve the document. It is possible that the document has been migrated back to primary storage from secondary storage, and therefore is now "present" (although possibly for only a limited time). Generally available immediately. Opposite of *migrated.*

Primary Storage

Storage area for electronic documents where the document is immediately available, without restriction. A document that has **NOT** been migrated is considered to reside in primary storage. In real life, primary storage should typically be the central magnetic (i.e. read-writable) disk. Contrasted with External storage, and secondary storage.

PSEP

Primary Storage Expiration Period (PSEP). Expressed in days. The period of time that a document should be stored in primary storage after reverse migration from secondary storage.

Purge

To delete a document or version of a document from the filing system. Bears no relation to DESTROY, which means a deletion due to the application of the schedule. A purge may be executed by the Information Manager at any time, regardless of the document's status vis a vis the schedule of the file containing it. Function for the Information Manager's use only.

Reference Date

File date of the earliest document filed.

Retention Period

Sum of active retention period plus dormant retention period, expressed in months and years.

Retrieve

To remove a copy of an electronic document from primary storage to one's personal working area, usually for printing or revision. The Information Manager my migrate documents from secondary to primary storage, so a user can retrieve a copy of it from primary storage.

R&D

Retention and Disposition, scheduling, schedules.

Secondary Storage

Place where non-disposition migrated documents are stored. Documents residing on secondary storage are generally less available than those residing on primary storage.

Section

A range of primary subjects each with unique numbers, titles and descriptions.

State

Refers to the state of either a volume or document sequence. Either Active, or Dormant.

Subject

A short description of the of contents of a group of related files and sub-files.

Target Date — Time frame of documents that have been requested for BF or charge-out. Can be a specific document date.

Universal Document Format A single, uniform format for the description and recording of electronic documents. e.g. ODA/ODIF. Non-proprietary. Defines all formatting information, and other support functions. Contrasts with Native Document Format.

Volume End Date Date of last correspondence in a volume, or date on which the volume ends.

Volume Start Date Date of first correspondence in the volume, or date on which the volume begins.

Volume Date Span Period covered from volume start date to volume end date, inclusive.

Volume Series of non-electronic documents i.e. does **NOT** refer to electronic documents stored and maintained by the CIMA.

REFERENCES AND FURTHER READING

"A National Archives Strategy for the Development and Implementation of Standards for the Creation, Transfer, Access and Long-Term Storage of Electronic Records of the Federal Government", National Archives, Washington, 1990.

Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records, National Institute of Standards and Technology (NIST), Washington, 1989.

Management of Electronic Records: Issues and Guidelines, Advisory Committee for the Co-ordination of Information Systems (ACCIS), United Nations, New York, 1990.

Managing Information in Office Systems: Final Report on the FOREMOST Project, National Archives of Canada, Ottawa, 1990.

Recommendations for a Compound Document Representation Format for the Canadian Patent Office, Strategic Technologies Inc., February 1989.

"Taking a Byte out of History: The Archival Preservation of Federal Computer Records", Twenty-fifth Report by the Committee on Government Operations, 101st Congress, 2nd Session, Washington, 1990.

"The Effects of Electronic Recordkeeping on the Historical Records of the U.S. Government", National Archives and Records Administration, National Academy of Public Administration, Washington, 1989.

"The National Archives and Office Systems: A Status Report", Records Management Journal, Volume 1, number 4.