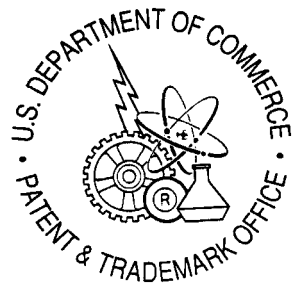
Prepared for:



LONG-TERM ACCESS and MIGRATION STRATEGY

FOR THE LIFE CYCLE MANAGEMENT OF ELECTRONIC PATENT AND TRADEMARK CASE FILES

February 26, 1999

Contract Number: 50-PAPT-700041 **Task Number**: 56-PAPT-8-05089

Deliverable: 98-03-10

Government Task Managers: Arthur F. Purcell, (703) 308-6868, FAX (703) 308-6916
Kathy Schultz, (703) 308-7400, FAX (703) 308-7407

Contractor: Cohasset Associates, Inc.
3806 Lake Point Tower
505 North Lake Shore Drive
Chicago, IL 60611
Tel. 312/527-1550

Contractor Task Manager: Richard D. Fisher, (408) 741-1287, FAX (408) 867-1289
e-mail: rd.fisher@worldnet.att.net

FINAL

Table of Contents

1 INTRODUCTION

Between now and the year 2003, the PTO is undertaking electronic patent and trademark filing, processing and maintenance programs that are part of a broader program of PTO modernization.

The electronic patent and trademark business processes can be broadly divided into an applicant component, and a PTO case file processing and maintenance component. At the point when a patent or trademark application or associated material is received from an applicant, or is created as part of the examination process by the PTO, it becomes a record copy and, therefore, is subject to the United States statutes and regulations governing the creation, use, preservation, and disposition of federal records.

Maintaining the *integrity* of patent and trademark records is required by law (18 U.S.C. § 2071. **Concealment, removal, or mutilation generally**). The integrity of the records in a patent or trademark case file must be maintained for their full retention life.

There are three requirements in 36 CFR that have particular relevance for the life cycle management of patent and trademark case files.

First, 36 CFR § 1234.22, Creation and Use of Text Documents**,** requires that the integrity of electronic records be maintained:

> (a) Electronic records systems that maintain the official file copy of text documents on electronic media shall meet the following minimum requirements:
>
> . . .
>
> (2) Provide an appropriate level of security to ensure integrity of the documents;

The second requirement in 36 CFR is to maintain the patent and trademark case files in a usable form and to ensure that they are "protected from unauthorized change or deletion for the length of their scheduled retention period" (36C FR §1222.50[8], (36 CFR §1234.30 [e], and 36 CFR §1234/32[c]). CFR §1234.30 Selection and Maintenance of Electronic Records Storage Media requires agencies to:

> (3) Retain the records in a usable format until their authorized disposition date; and (4) if the media contains permanent records and does not meet the requirements for transferring permanent records to NARA . . . , permit the migration of permanent records at the transfer to a media which does meet the requirements." (36 CFR Ch. XII, Subpart C, 1234.30 [7-1-98 Edition], p. 613)

Although this section of the CFR does not explicitly define the meaning of "usable format", other CFR references indicate that the overall intention is to ensure long-term access to intelligible electronic records that have been protected from alterations and deletions or loss resulting from changing technology or deterioration. (36 CFR §1234.30(e)). Access, it should be noted, not only includes indexing tools and the like that support search and retrieval, but also encompasses all of the technical and intellectual requirements that are necessary for electronic patent and trademark case files to be fully usable and understandable during their full life cycle.

Third, under 36 CFR §1228, Disposition of Federal Records, the PTO is required to develop a records schedule, which the Archivist of the United States must approve, that establishes a retention period for all records, including patent and trademark case files (36 CFR §1234.34).

The PTO 1997 Comprehensive Records Schedule identifies the retention period for issued patents, and abandoned patents referenced by an issued patent, as permanent. The PTO has custody and responsibility for the case files of issued patents for forty years, after which they are transferred to the National Archives and Records Administration (NARA) for permanent retention. Registered trademark case files are retained by the PTO until they are cancelled or expire. Trademark case files selected for permanent retention are transferred from the PTO to NARA six years after they are cancelled or expire, otherwise the non-selected case files are retained by the PTO for two years prior to disposal.

It is highly likely that over the next four decades or so there will be enormous changes in digital storage and processing technologies and, therefore, the PTO should have in place a program that supports long-term access to electronic patent and trademark case files, no matter what changes occur. It is for this reason that this study on a long-term access and migration strategy for electronic patent and trademark case files was undertaken.

Section 1 of this report contains introductory and background information about the document. Section 2 defines the purpose and scope of the deliverable. Section 3 contains a Management

Summary that highlights salient features of the report. Section 4 reports the findings of several recently completed research projects and reports that address issues associated with long-term access to electronic patent and trademark case files. Section 5 reviews the conceptual foundations underlying the study, including recordkeeping application environment, technology obsolescence, archiving, authenticity and reliability, media renewal, transfer, migration, and migration versus transfer. Section 6 articulates a long-term access strategy for the PTO to ensure that electronic patent and trademark case files remain usable and accessible throughout their authorized disposition date, regardless of changes that may occur in information technologies. Section 7 contains recommendations for a long-term access and migration strategy for the PTO to consider in the implementation of the information infrastructure and application information systems (AIS) for electronic patent and trademark case files. These recommendations are based upon a synthesis of conceptual foundations discussed in Section 5 and electronic patent and trademark case file preservation strategy discussed in Section 6.

2 PURPOSE AND SCOPE

The purpose of this report is to set forth a long-term access and migration strategy that:

- Provides long-term access to electronic patent and trademark case files during their authorized retention period,
- Satisfies legal and regulatory requirements, and
- Can be incorporated into the life cycle management of electronic patent and trademark case files.

The concept of long-term access includes three fundamental aspects of electronic records management over the full life cycle:

1) Preserving the integrity, and, as such, the trustworthiness of electronic case file records,
2) Ensuring the processibility of electronic case file records through the capability to automatically access and render electronic case file records with their original structure, content, and context, and
3) Providing for the transferability of electronic case file records without any loss in structure, content, or context from one technology environment to another that result from major upgrades or renewals of system hardware or electronic recordkeeping functions.

The term migration is used in this report as an umbrella concept that emphasizes ensuring the processibility and integrity of electronic patent and trademark case files regardless of the specific technology platform in use. The specific methods for conducting migration, which include media renewal (copy and reformat) and transfer, take into account variations the potential effects of migration on electronic patent and trademark case files, and the requirements for providing controls and documentation of these migration methods.

3 MANAGEMENT SUMMARY

Between now and the year 2003, the PTO is undertaking an electronic patent and trademark filing, processing and maintenance program that is part of a broader goal for PTO automation. The inherent result of the electronic workplace is the receipt and creation of records in electronic format. At the point when a electronic patent or trademark application or associated material is received by the PTO or is created as part of the examining process by the PTO and committed to a case file, it becomes subject to Code of Federal Regulations (CFR) governing the creation, use, preservation, and disposition of federal records.

Federal regulations require agencies to:

> (3) Retain the records in a usable format until their authorized disposition date; and (4) if the media contains permanent records and does not meet the requirements for transferring permanent records to NARA . . . , permit the migration of permanent records at the transfer to a media which does meet the requirements." (36 CFR Ch. XII, Subpart C, 1234.30 [7-1-98 Edition], p. 613)

The PTO 1997 Comprehensive Records Schedule identifies the retention period for issued patents, and abandoned patents referenced by an issued patent, as permanent. The PTO has custody and responsibility for the case files of issued patents for forty years, after which they are transferred to the National Archives and Records Administration (NARA) for permanent retention. Registered trademark case files are retained by the PTO until they are cancelled or expire. Registered trademark case files selected for permanent retention are transferred from the PTO to NARA six years after they are cancelled or expire, otherwise the non-selected case files are retained by the PTO for two years prior to disposal. The PTO will have to retain many of the electronic patent and trademark case files and records in accessible and usable form for 40 years of longer.

Over the next four decades or so there will be enormous changes in digital storage and processing technologies and therefore the PTO must have in place a strategy and program that

supports long-term access to and preservation of electronic patent and trademark case files no matter what changes occur in digital storage and processing technologies.

3.1 BACKGROUND

The PTO has many years of experience in managing digitally stored files consisting of full-text and digitized, compressed images of published patents, as well as management of the PALM database (bibliographic, retrieval and tracking metadata) that is used to track and provide status information on pending and issued patents and abandoned patent applications. The PTO has recently undertaken a major media renewal by copying compressed, digitized image files of published patents from optical media to magnetic media. The final stage of this migration will be a transfer and reformatting of published patent images, which are in a proprietary image and compression format, to a more open image format (TIFF) and an accessibility environment that is designed to better support long-term access.

However, based on the definition of a Federal record (44 U.S.C. 3301), published patent databases (image and full text) are considered to be non-records or temporary files, not the legal record copy. The legal record copy, namely the patent and trademark case files, have always been managed in paper form.

Conceptually, there are similarities between the management of "temporary" digital patent databases and the electronic record copies that will reside in case files. However, there are a number of unique considerations related to the management of electronic records that are driven primarily by the requirement to preserve the integrity and, as such, the authenticity and reliability of the electronic legal record copy. This is complicated by the requirement for the PTO to retain custody of issued patent case files for forty years and registered trademark case files for a potentially indefinite period of time. These unique areas are discussed as part of the following conceptual foundation.

3.2 CONCEPTUAL FOUNDATION

While this report covers the strategies for migration of electronic records from one technology or technical solution to another, its overall focus is on the broader strategy and requirements that ensure the preservation of authenticity and reliability of electronic records over the full record disposition period. There are five requirement areas that need to be addressed in a long-term access strategy.

The first area is fundamental and addresses the integrity and, as such, the trustworthiness of electronic patent and trademark case files over the full retention period, namely that the electronic records remain unaltered. This means that from the time any type of electronic patent or trademark record, whether received from an applicant or created by the PTO, is committed to the case file, its structure, content, and context cannot be changed by anyone, including the individual or organizational component that initially created, used, and maintained it.

The second area involves ensuring the processibility and transferability of electronic case files over the full retention period. Processible means that the electronic case files can be read, accurately interpreted, and rendered on a monitor or by a printer with current computer technology without regard for the technology environment in which they were created, used and maintained. Transferable means that the case files can be moved from one automated information system (AIS) environment (e.g., electronic recordkeeping functions) to another with no material or visible loss in content, structure or context.

The third area addresses preserving the accessibility of electronic patent and trademark case files using recordkeeping functionality that includes: preserving the integrity of electronic records; maintaining the readability; ensuring their intelligibility, reconstructability, understandability, and encapsulating with them appropriate metadata. These functions are intended to mitigate any diminution in the legal status of electronic patent and trademark case files as authentic and reliable records, despite changes in computer technology. Whether these functions are provided as part of an AIS or in a separate electronic recordkeeping system, the care and handling of electronic records should be guided by the policies and procedures of the PTO Records Management program.

The fourth area deals with the development and implementation of rigorous policies, procedures that carefully describe the circumstances under which actions are taken to ensure the processibility and transferability of electronic patent and trademark case files, including the identification of the organizational unit, individual(s), and automated tools that were involved.

The fifth area relates to the use of non-proprietary information technology standards, especially those dealing with file formats. (For more information on file formats see the report deliverable 98-03-08 for Task Order 98-03. *Assessment of File Formats to Support the Life Cycle Management of Patent and Trademark Case Files*).

The considerations and alternatives for addressing these areas and the recommendations regarding appropriate solutions for the needs of the PTO are set forth below.

3.3 STRATEGY AND RECOMMENDATIONS

The strategies and recommendations are designed to ensure long-term access to electronic patent and trademark case files and cover the following areas:

- The importance of preserving the authenticity and reliability.
- The importance of a clear policy statement regarding long-term access.
- The need for clearly defined quality control procedures.
- The actual practices that the PTO should follow in providing long-term access to reliable patent and trademark case files.
- The importance of using non-proprietary standards for facilitating long-term access.

3.4 GENERAL POLICY

The mandate of the PTO to provide long-term access to usable electronic patent and trademark case files should be explicitly reflected in a policy statement. A sound policy statement should specify the objectives of the program and the circumstances under which media renewal (reformatting, and copying) and transfer activities are initiated, and the types of compliance audits that will take place.

3.5 QUALITY CONTROL

Archival science and the Federal Rules of Evidence give the greatest weight of trustworthiness to records (patent and trademark case files) that are created and used in accordance with established controls and procedures. Whatever actions the PTO takes to ensure long-term access to electronic patent and trademark case files should be carried out in accordance with written procedures.

Among the quality control techniques to be considered are comparisons of Cyclical Redundancy Checksums (CRCs), hash digests, and before and after bit/byte comparisons that identify whether any information has been lost during media renewal or transfer. To ensure compliance with these procedures, periodic quality control audits should be conducted.

3.6 ELECTRONIC RECORDKEEPING FUNCTIONS

The following functional areas of electronic recordkeeping are necessary to provide for long-term access and preservation of electronic patent and trademark case files:

- *Unaltered*: Under Rule 901 of the Federal Rules of Evidence, authentic electronic patent and trademark case files means that they are what they purport to be -- that is, trustworthy case files whose structure, content, and context have not been altered, changed, or otherwise corrupted.
- *Readable*: The bit stream comprising patent and trademark case files can be processed on the computer system or device that initially created it, the computer system or device that currently stores it, or the computer system or device that will be used to store the digital information in the future.
- *Intelligible*: A function of information about what the bit stream in fact represents and the operating software's capacity to take appropriate action based upon this information.
- *Understandable*: Electronic patent and trademark case files are understandable because they are intelligible to both a computer and humans. Electronic material is intelligible to humans when it conveys information about a person, an event, concept, person, or related phenomenon.
- *Reconstructible*: Electronic records can be rendered so that they have the "same logical and physical structure and intellectual content" as when initially moved into an electronic recordkeeping application environment.

- *Encapsulated*: All of the information associated with a specific electronic patent case file, such as metadata, content, and linkages to multimedia, exists as a single logical or physical entity.

Satisfying these recordkeeping functions could be attained as an integral part of one or more primary patent or trademark AISs. However, considering the forty year period that issued electronic patent case file records are retained by the PTO, and the indefinite period of time that registered electronic trademark case files could be retained, a separate but accessible electronic recordkeeping system could potentially provide a more cost effective means of supporting these preservation requirements over the long term.

It is recommended that electronic recordkeeping functions be developed, either as an integral part of an AIS system or provided as part of stand alone recordkeeping system, that meet the following requirements:

- Prevention of any changes to the structure, content, and context of electronic patent and trademark case files using access controls, read-only viewing controls and version control.
- Retrieval of electronic case files based upon a data elements ("properties") in profile metadata or full-text search.
- Encapsulation of all components of an electronic patent or trademark case file, including content, metadata for long-term retention and metadata of results from quality controls applied during media renewal or transfer of electronic records.
- Capture and preservation of all long-term retention metadata identified in Task Order 98-Deliverable 03-11, *Metadata Requirements for Long-term Access and Retention of Electronic Patent and Trademark Case Files*.
- Expansion of profile metadata to include relevant information useful for preservation purposes, such as the history use file (below)
- A use history file that contains specified metadata (e.g., CRCs or hash digests) generated during the reformatting, copying, or transferring electronic patent and trademark case files
- Capture and preserve as metadata or as audit trails the quality control decisions and the results of actions taken to ensure long-term access to reliable electronic patent and trademark case files.

3.7 MIGRATION: REFORMAT, COPY, AND TRANSFER

In order to address this requirement, a long-term access strategy is proposed that includes a component to ensure the processibility of electronic patent and trademark case files in a current electronic recordkeeping environment and their transferability from this dependency application environment to a new one as the PTO upgrades its Information Technology infrastructure or installs a completely new technology platform. The specific means proposed for achieving these objectives are media renewal, which consists of reformatting and copying electronic patent and trademark case files, and transfer, which consists of moving electronic patent and trademark case files from an existing technology platform to a new one.

The recommendations are to develop and implement procedures for reformatting, copying, and transferring electronic patent and trademark case files that include the following:

- Copy and reformat, if required, electronic case file records at the time they are moved from one electronic recordkeeping application environment to another.
- Reformat electronic case files when new storage devices or media are utilized.
- Copy electronic case files every ten years (36 CFR 123430[g][5]) if magnetic tape is the storage medium.
- Copy electronic case files when the annual readability sample discloses ten or more temporary or read “errors” on a specific storage medium
- Transfer electronic case files to a technology platform when the current software is upgraded or a new electronic records management system is installed
- Ensure the trustworthiness and integrity of reformatted, copied, and transferred electronic case files by employing a strict quality control procedure that may include bit/byte comparisons and comparisons of hash digests and CRCs.
- Document fully all actions taken when reformatting, copying and transferring electronic case files and include this information in the metadata associated with each case file
- Provide for disaster recovery by creating two copies of electronic case files at the time of reformatting, copying, or transfer and store one copy as a vital record backup at an off-site location.

3.8 PHYSICAL SECURITY

Protecting the physical security of electronic patent and trademark case files from alteration, change, or loss should be a critical component of the mission of a PTO storage repository in which electronic patent and trademark case files are stored. The following recommendations are offered for achieving this objective:

- Maintaining a controlled access storage facility and monitor compliance with access control procedures.
- Locating the storage facility in an area where the threat of a natural disaster is minimal.
- Creating a disaster vital record backup copy that is stored at a separate physical location and developing and testing a disaster recovery plan.
- Adhering to the recommendations of *NBS Special Publication 500-101 Care and Handling of Computer Magnetic Storage Media* (pp. 37 - 52) regarding exposure to magnetic fields if a magnetic storage medium is used.

3.9 ENVIRONMENTAL CONTROL AND MONITORING

The continued readability of electronic patent and trademark case files can be establishing by developing a program that provides: 1) a stable storage environment, and 2) proper care and handling procedures. Such a program should include the following:

- Implementing a program based upon 36 CFR 1234.309(g)(4) to read annually a statistical sample of the storage media to identify real or impending loss of information
- Maintaining a stable storage environment in which the temperature and relative humidity (RH) are 15º C (59° F±5) and 40% (±5)
- Installing a filter system to remove air dust particles and gas pollutants
- Prohibiting the consumption of food and beverages and smoking in the storage facility.

4 RESEARCH PROJECTS AND OTHER STUDIES

A number of research projects and studies have examined electronic records issues and requirements from a theoretical, policy, and implementation perspective. Each of these studies has made a significant contribution to the state of archival knowledge and understanding of critical issues that electronic records pose. This section of the report reviews nine research projects and studies that contain useful background information for the PTO life cycle management of electronic patent and trademark case files.

4.1 UNIVERSITY OF PITTSBURGH

Between 1993 and 1996 James Williams and Richard Cox (along with David Bearman and other associates) conducted a study to identify the “functional requirements” for electronic records management, one of which is migration. In addressing migration, the project team focused on two requisites that are essential in the design of systems for the creation and management of electronic records: “First, the assurance that such records satisfy the requirements for evidence and second, the methods by which records can be made available over time without constant re-presentation and migration of their intellectual contents." (David Bearman, 'The Physical Archives and the Virtual Archives, " *Archivum* XLIII (1997): 1.)

The research associated with assuring that electronic records satisfy evidentiary requirements focused on practical considerations, beginning with what is called the “literary warrant” for records. A literary warrant is a statutory, regulatory, or organizational policy requirement to create and maintain records as evidence. A review of more than two hundred such “literary warrants” identified core record requirements that formed the basis of fourteen functional requirements. Of particular interest to this study is a technical report by David Bearman and Ken Sochats on the production rules and metadata requirements for business-acceptable communications as evidence that can also support the “periodic migration” of electronic records. The metadata requirements for business acceptable communications are divided into six layers:

Handle, Terms and Condition, Structural, Context, Content, and Use History. Each of the six layers is then subdivided into a number of data elements that flesh out the requirements. The sixth level -- Use History -- identifies metadata that tracks (i.e., a usage audit trail) the use history of each record, from creation through various stages of processing, which has a direct bearing on ensuring the trustworthiness of electronic records.

4.2 UNIVERSITY OF BRITISH COLUMBIA

The University of British Columbia Project, which is entitled "The Preservation of the Integrity of Electronic Records," offers a largely theoretical perspective. Project Director Luciana Duranti and her colleagues, Terry Eastwood and Heather MacNeil, focused upon the concept of a record, the nature of a record, and the identification of a record's component parts by using an inductive methodology that takes into account fundamental concepts of archival science that have been known in Europe for several centuries.

One of the major conclusions of the UBC project is a clear and explicit articulation of the concepts of reliable and authentic electronic records. The reliability of records involves their trustworthiness as "proof and memory of the activity of which they constitute the natural by-product, that is, to their ability to stand for the facts they are about." The capacity of records to "stand for the facts they are about" is determined by procedures that control their form and creation and the trustworthiness of the creator in ensuring their accuracy.

Equally important is the authenticity of electronic records, which the UBC project defines as "reliability over time."[1] From this perspective, ensuring the authenticity of electronic records requires protecting records against alteration from the moment of transmission through any reproduction and preservation procedure. This is crucial because providing long-term access to electronic records involves operations and procedures that expose the records to the risk of alteration or loss of contextual information. The authenticity of electronic records can be ensured

[1] It should be noted that in the Federal Rules of Evidence the meaning of authenticity is functionally the same as the archival science definition of reliability and the meaning of reliability is functionally the same as the archival science definition of authenticity.

only in the form of reliable copies, i.e., records whose preservation is made possible through an uninterrupted sequence of reproduction that is done in accordance with rigorously controlled procedures.

4.3 US DEPARTMENT OF DEFENSE RECORDS MANAGEMENT PROJECT

In 1993 the U.S. Department of Defense established a Business Process Reengineering Task Force to identify opportunities to improve the records management process, especially in light of the increasing use of distributed network technologies to create and use records. Two years later a Records Management Task Force was created to address electronic records management issues. Working with the UBC project staff, the task force used the Integrated Definition Language (IDEF) to develop a business entity and process model for managing electronic records. Based upon these two models, the DOD project staff issued several reports, including a Design Criteria Standard for Electronic Records Management Software Applications (DOD 5015.2-STD) that was issued in November 1997. The standard identifies "mandatory and optional baseline functional requirements for Records Management Applications (RMA) software" that are used to certify RMA software that complies with the standard. Among the mandatory requirements are the capacities to:

- manage records without regard for storage media or other characteristics
- implement file plans and allow users to assign a file code to each record
- prevent changes to documents that have been designated as records
- permit creation of profile data (metadata) when a record is filed
- designate a record as a vital record
- date a record automatically when it was saved as a record
- treat electronic mail messages as records
- assign disposition schedules
- retrieve records based upon a specified list of data elements

- delete records and records profiles scheduled for destruction
- provide audit utilities of records capture, retrieval, and preservation activities to ensure their reliability and authenticity
- read and process records

Currently, three software applications – TRIM from Tower Software, FOREMOST from Provenance Systems, and PC DOCS from PCDOCS Corporation -- have been certified as being in compliance with these functional requirements.

The language used to describe the "read and process records" functional requirement merits quoting below because with the exception of the somewhat ambiguous meanings for the terms: *format of stored records*, *reliability of the record*, and *migration*, it tracks closely with the focus of this study.

> Since RMAs [Records Management Applications software] are prohibited . . . from altering the format of stored records, the organization shall ensure that it has the ability to view, copy, print, and if appropriate, process any records stored in RMAs as long as that record must be retained. The organization may meet this requirement by maintaining the hardware and/or software used to created or capture the record; by maintaining hardware and/or software in its native format; by ensuring downward compatibility when hardware and/or software is updated; or by migrating the record to a new format before the old format becomes obsolete. Any migration shall be controlled to ensure continued reliability of the record.

4.4 PRESERVING DIGITAL INFORMATION: RLG/CPA TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION

Another widely publicized investigation into the problem of long-term access to electronic records is the work of a task force on preserving digital information sponsored by the Research Libraries Group (RLG) and the Commission on Preservation and Access (CPA).[2] In January

[2] The functions of the Commission on Preservation and Access have been incorporated into the Council on Library Resources (CLR).

1995 RLG and the CPA established a "Task Force on Archiving of Digital Information" and charged it with the task of examining the problems and issues associated with providing "continued access indefinitely into the future of records stored in digital electronic form." The focus of this task force was on the impact of digital technology obsolescence on libraries and archives. The Task Force report was published in May of 1996.

The Task Force report contains a cogent presentation of the limitations of digital technology, particularly with regard to technology obsolescence. The report argues that the present approach to technology obsolescence through "periodic refreshing of digital information" by copying it to new media fails to cover the full range of issues. Migration, the report holds, is a much broader term that encompasses a number of organized tasks and options. Migration is defined in the report as:

> the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of a constantly changing technology

The report includes a detailed cost-benefit analysis of the use of digital technologies for preserving books and library materials vis-à-vis the cost of preserving the same books in paper form for ten years. The report concludes with nine recommendations for follow-up action by the Commission on Preservation and Access and the Research Libraries Group that relate to funding cooperative retrospective digital conversion services, demonstration projects, and practical experiments in the archival applications of a variety of technology tools.

4.5 SESAM

Another project is the work of Ulf Andersson, an information technology specialist at the ASTRA Corporation, an international pharmaceutical company in Sweden. Andersson's report looks at the migration of electronic records from a business information model perspective. Andersson employs concepts and technical language that are unfamiliar to many archivists and records managers, especially those in North America, and consequently the archives and records management communities may not fully comprehend the importance of the study. Nonetheless it

is a seminal study. It incorporates key concepts from archival science and advances our knowledge of technological solutions and techniques that are available today, are under development, or should be developed. In addition, it employs the concept of encapsulating electronic records in SGML described containers. Furthermore, the study suggests that the business community and archives can forge an alliance to promote the development and implementation of generic tools and standards that support the migration of reliable electronic records.

4.6 NORDIC COUNCIL FOR SCIENTIFIC INFORMATION

A closely related effort to address the issue of long-term access to electronic records is a project by the Nordic Council for Scientific Information that resulted in the publication of a report entitled *To Preserve and Provide Access to Electronic Records*. The Nordic Council study focused on three key issues: 1) development of software for transferring database records with a flat file structure to a relational database structure in order to facilitate on-line access, 2) investigation of adequate storage media for long-term storage of electronic records, and 3) analysis of different physical arrangements for long-term storage of electronic records.

This report introduces several important considerations regarding long-term access to reliable electronic records. From the Nordic perspective, the long-term storage format for database records is a flat file structure. On-line access to these records requires that they be converted to a SQL-conforming database application. The Nordic report also addresses the question of the appropriate long-term storage media for electronic records and identifies eleven criteria that should be kept in mind when selecting media for the long-term storage of electronic records. Based upon these criteria, Digital Linear Tape (DLT) and 4 mm and 8 mm tape were identified as the most suitable for long-term storage of electronic records in the Nordic context.

4.7 INTERNATIONAL COUNCIL ON ARCHIVES

The Electronic Records Management Committee of the International Council on Archives (ICA) initiated a project in 1994 that culminated in the release of three studies in June 1996. The first

study is a thorough and persuasive guide that is "designed to help archival institutions reposition themselves to address the management of electronic records." It is divided into two parts, the first of which examines the impact of technological, organizational, and legal trends on institutions, particularly archives, to keep and manage records in electronic form. Based upon this examination, the study proposes a number of strategies that archival institutions should consider. The second part focuses on tactics that archival institutions may wish to adopt. The Committee anticipates further work over the next year or so in fleshing out these "how to" techniques. The second study is a survey of current programs to manage electronic records that are planned in 134 archival institutions around the world. The third study is a literature review dealing with electronic records prepared for the committee by Alf Erlandsson, former Archivist of the United Nations. These three studies collectively are a valuable and convenient general resource for archivists and records managers. However, the studies do not offer very many specific guidelines or recommend best practices.

4.8 OPEN ARCHIVAL INFORMATION SYSTEMS (OAIS) PROJECT

In 1995 Subcommittee 13, Space Data and Information Transfer Systems, Technical Committee (TC) 20, Aircraft and Space Vehicles of the International Standards Organization requested that its working body, the Consultative Committee for Space Data Systems (CCSDS), develop an archive standard to help preserve data derived from space investigations. The CCSDS responded with a proposal to define an archives reference model that was intended, among other things, to address intermediate and indefinite long-term storage of digital data. This proposal was precipitated in part by the growing needs for long-term preservation of digital data from space investigations, an incomplete understanding of digital preservation issues, and constant changes in technology. Although the impetus for this proposal came from the space data community, early on it was recognized that the effort cut across all communities concerned with long-term access to digital material. Consequently, participation was broadened to include social science, arts, and humanities data libraries, libraries, archives, and museums.

In June 1998 a conference sponsored by the National Aeronautics and Space Administration, CCSDS, Johns Hopkins University Applied Physics Laboratory, Committee on Earth Observation Satellites, and Research Libraries Group was held at the United States National Archives and Records Administration to further expose the OAIS model and to determine interest in the development of additional broadly based archival standards or best practices. The primary purpose of the conference was to clarify language and concepts so as to make the model more understandable and useful to a very broad archival community as the reference model moves through the international standards arena with the expectation of gaining CCSDS and ISO approval by December 1999.

In September 1998 version 4 of the OAIS Reference Model (available at http://ssdoo.gsfc.nasa.gov/nost/isoas/us/overview.html) was released. The OAIS Reference Model incorporates a functional model and a logical model for archival storage that addresses most of the major components of archival storage, albeit in language and terminology unfamiliar to many archivists and records managers. For example, the model uses "ingest" to describe what archivists and records managers would call acquisition. Chapter 5 is entitled "Digital Migration Perspectives" and it includes a definition of digital migration and a discussion of specific methodologies that can be used to ensure long-term access to spatial data. Digital migration is defined as the "transfer of digital information while intending to preserve it, within the OAIS. " The report identifies four primary instances of digital migration: Refreshment, Replication, Repackaging, and Transformation. Of these four instances of digital migration the only one that results in content change of data is transformation in which the objective is to update the digital data with more accurate or reliable digital data. The result of digital migration transformation is a new version of the digital data and any associated information.

The OAIS Reference Model is intended to serve the needs of the space data community so it reflects specific concerns that are not easily generalizable to the needs of the PTO without some modification in terminology and concepts. The OAIS Reference Model correctly identifies the transformed scientific data as a "new version." However, electronic patent and trademark case files must remain unrevisable so transformation as it is defined would be inappropriate.

4.9 DLM FORUM

In December 1996 the European Commission and the Member States of the European Union organized a forum that brought together experts from industry, research, public administration, and archives to discuss archives services as an essential component of an information society. One purpose of the forum, known as the DLM-Forum, was to review a set of multidisciplinary guidelines on best practices to follow in defining short and medium-term strategies of managing and preserving electronic records. In 1997 the DLM-Forum published *Guidelines on Best Practices for Using Electronic Information: How to Deal with Machine-Readable Data and Electronic Documents* (Updated and Enlarged Edition). The guidelines offer very useful information regarding the creation and maintenance of electronic material, the conversion of paper documents to electronic form, and the selection of file formats. Unfortunately, the only guidance regarding long-term access to electronic records is their conversion to a standard format. In addition, the guidelines treat migration and conversion as synonymous terms.

5 CONCEPTUAL FOUNDATIONS

A key factor in the development of a comprehensive long-term access and migration strategy for electronic patent and trademark case files is a sound understanding of the key issues and possible solutions. Many archivists and records managers tend to view long-term access to electronic records largely as a technical issue requiring technical solutions, when in fact there are conceptual or definitional issues that must be clearly understood before moving to technical solutions. This is particularly evident in the way certain concepts (e.g.,"conversion or migration" in 36 CFR §1234.22(3)) are synonymous or interchangeable without taking into account that there may be significant procedural or technical considerations that make them distinctively different. As such, this section of the report discusses a number of key concepts or terms in order to foster a clear understanding of the problems of providing long-term access to electronic patent and trademark case files.

5.1 PAPER RECORDS, DATA FILES, AND ELECTRONIC RECORDS

A fundamental premise underlying this report is that the issues and techniques involved with managing paper records and data files are substantially different than those involved with managing electronic records. What may be suitable for paper records or data files is not suitable for electronic records.

Paper Records. Paper records exist as physical entities whose structure, content, and context are inseparably linked together. This inseparable linkage is manifested in a physical format (text, image, or a spreadsheet), a numerical or color coded filing scheme, and physical placement in a folder with other related records, to name only a few. In contrast, data files are logical entities consisting of binary data (1s and 0s) whose structure and content may exist independently of each other.

Data Files. Like paper records, data files have structure and content. However, the structure of data files is limited to a tabular format and a data dictionary defines the content of a specific table or combinations of tables. In order for the structure and content of a data file to be manifested as a coherent table or combination of tables, software must interpret the binary data, reconstruct the parts, and render the record for display on a monitor or printer.

Electronic Records. Electronic records and data files share in common the fact that they exist as logical entities consisting of binary data (1s and 0s) whose structure and content may exist independently of each other and are linked together for rendering by software. However, the structure of electronic records is much more complex and rich in that a variety of formats - text, image, vector drawings, spreadsheets, tables, audio, and moving images - may exist in a single record or as a linkage of files that comprise a record. In addition, electronic records have a context of maintenance and use that is essential to being admissible as evidence and, as such, whose integrity must be preserved for their full retention life. Consequently, the life cycle management of electronic records is a very complex task involving a number of critical issues and techniques that are not associated with the management of either data files or paper records.

5.2 AUTHENTICITY AND RELIABILITY

What constitutes authentic and reliable electronic records? Federal Rules of Evidence 901 declares that an authentic document submitted for admission as evidence is one that is what it purports to be. Consequently, the trustworthiness of electronic records (which applies as well to paper records) is derived from their ability to stand for the facts they are about. Authenticity is derived initially from the context of its creation, storage and use.

Reliable electronic records are those records that retain their authenticity over time. "Over time" encompasses, among other things, the circumstances of the initial transmission of electronic record as well as preserving the integrity of the record during the period of use and maintenance, including migration of the record for media renewal or transfer to another system hardware or software or recordkeeping application environment.

Electronic patent and trademark case files, of course, will not exist as physical entities. Preserving their integrity will be particularly challenging because digital information technologies make it relatively easy for the context of use to be altered, either by accident or by design, without leaving any visible evidence. With processible text, digits can be transposed or the word in a sentence could be modified slightly to give an entirely different meaning. Digital images can be cut and pasted to create such realistic images that even experts may not be able to detect the difference. Techniques and approaches that help protect the integrity of electronic patent and trademark case files are discussed in Section 6.1.

5.3 ELECTRONIC RECORDKEEPING FUNCTIONS

The primary purpose of an electronic recordkeeping functions is to support storing, accessing, and preserving records as evidence that can be used in carrying out legal, operational, and historical activities. An electronic recordkeeping system or its equivalent functionality should be distinguished by the fact that its contents have been retained as unrevisable evidence of transactions as required for statutory and regulatory compliance and accepted business practices.

Satisfying these recordkeeping functions could be attained as an integral part of one or more primary patent processing AISs. However, considering the very long year period of time (forty years for issued patent case files, and indefinite for registered trademark case files) that electronic records will be in the legal custody of the PTO, a separate but integrally accessible electronic recordkeeping system could potentially provide a more cost effective means of supporting these preservation requirements over the long term.

These recordkeeping functions are designed to protect case files and any associated records from loss, alteration, and deterioration so that they may be accessed and used as trustworthy evidence as far into the future as is necessary. To fully achieve these objectives these recordkeeping functions must provide for the preservation of the integrity of electronic case files so that over time they are readable, intelligible, understandable, reconstructible, unalterable, and encapsulatable. This means, among other things, that the AIS or external utilities must capture metadata such as hash digests, CRCs, and the results of media renewal and transfer for integrity audit purposes.

Because records in these patent and trademark case files would have certain features, such as non-proprietary specific file formats (Task Order Deliverable 98-03-08 *Assessment of File Formats to Support the Life Cycle Management of Patent and Trademark Case Files*), the electronic recordkeeping system must support all of the file formats used in the patent and trademark examination process. In addition, certain metadata that is captured by PALM, for instance, should be used to populate the metadata profiles supported in the electronic recordkeeping system and provision must be made to support this interface. It is important to remember that records in an electronic recordkeeping system may be required for evidence of transactions or may contain reliable information that is useful for other purposes such as historical analysis. Consequently, an electronic recordkeeping system should support powerful search and retrieval functions that would facilitate timely access.

5.4 TECHNOLOGY OBSOLESCENCE

Joseph P. Martino, a research scientist at the University of Dayton, has suggested that it is useful to differentiate between a technology and a technical approach. A technical approach, he writes is "a specific technical means of solving a problem or performing a particular function." For example, the open hearth and Bessemer processes are different technical approaches to the technology function of steel making just as vacuum tubes and transistors are different technical approaches to the technology function of rapid switching of electrical signals.

For the purposes of this study, technology obsolescence should be understood as resulting from major changes in technical approaches that supersede or displace established technical approaches. Data storage and retrieval are computer technology comprised of storage media, storage devices and storage systems software. Technology obsolescence occurs in storage media when newer and better storage media (e.g., *6250* bpi magnetic tape) displace older, established storage media (e.g., *1600* bpi magnetic tape). The storage function continues, but the technical solution changes. Technology obsolescence also occurs in operating systems and software applications as vendors introduce products with new functionalities. Many vendors attempt to mitigate technology obsolescence by providing "backward compatibility" over several generations of media and software..

There are three other dimensions of technology obsolescence that merit consideration. First, the history of digital technological innovation over the last four decades or so offers compelling evidence that technological change is inevitable and irreversible. Second, vendors tend to induce technology obsolescence through the introduction of new products or modifications of existing ones to achieve or maintain a competitive advantage in the marketplace. Third, marketplace displacement does not automatically mean technology obsolescence. Practically speaking, as long as current devices continue to function properly and spare parts and systems maintenance are available, then technology obsolescence has not occurred. Maximizing the life span of old systems, however, only delays the inevitable.

5.5 MIGRATION

The definition of migration used in *Preserving Digital Information* (Section 4.4 of this study), which essentially means ensuring that present and future users can retrieve, display, and otherwise use electronic materials in the face of constantly changing technology, is an improvement but is too broad. The definition includes copying and reformatting digital storage media, replacement of storage devices, and transfer of electronic records from one technology application to a newer one with and without loss of record integrity. The complexity of technical issues and the scope of human resources required to reformat, copy, convert, and migrate records differ substantially and therefore cannot be adequately conveyed by "migration."

Although the OAIS Reference Model definition of digital migration follows the same general direction as that offered in *Preserving Digital Information*, it makes much more explicit the specific circumstances in which use of each of the four types of digital migration (Section 4.8 of this study) is appropriate. Certainly, this is a noteworthy contribution and expands our knowledge and understanding of certain technical aspects of long-term access to electronic records. However, from the archival science and records management perspectives, the emphasis on the accuracy of space science digital data vis-à-vis its authenticity or trustworthiness is undesirable in programs providing long-term access to reliable electronic records.

While generally agreeing with the broader definitions set forth in these two studies, this report takes a more specific approach in terms of the actual functions required to carry out migration. The notion of a long-term access strategy is defined as two primary activities: (1) maintaining processible reliable electronic patent and trademark case files through media renewal - reformatting and copying - and (2) transferring reliable electronic patent and trademark case files from a current dependency software to a newer one with no loss in their structure, content, or context to a newer one as technology changes.

Transferring electronic patent and trademark case files shares with media renewal the requirements for readable media, operational devices to read the media, and target media and devices. It differs from media renewal in that a specific software application path such as a target technology platform with backward compatibility is required. As long as the PTO continues to support technology upgrades and the introduction of new technology platforms that support backward compatibility during the scheduled PTO retention of electronic patent and trademark case files, the mechanical aspects of transfer for electronic patent and trademark case files should be fairly straightforward.

5.5.1 Media Renewal - Reformat

As noted earlier, maintaining the processibility of reliable electronic patent and trademark case files means ensuring that they can be read and accurately interpreted by contemporary computers. Achieving this objective of media renewal involves reformatting and copying. The following section discusses reformatting and its role in media renewal.

The format of electronic patent and trademark case files (i.e., records) involves both the internal representational or logical structure of the 1s and 0s (e.g., ASCII, EBCDIC, Unicode, or binary) and their encoding for storage on a storage medium or transmission on a carrier. At its simplest level, reformatting of electronic patent and trademark case files occurs when there is an alteration in the underlying bit stream but there is no change in their representation or intellectual content. Typically, reformatting is associated with moving a bit stream from one storage medium to a different one. For example, moving electronic patent and trademark case files stored on 9 track 6250 bpi tapes to 18 track 3480 tapes constitutes reformatting because there is a change in the physical carrier even though when the reformatted electronic patent and trademark

case files are rendered for viewing there is no alteration in their appearance or content. Reformatting also can occur when there is a change in the character code that represents the 1s and 0s but there is no change apparent in the rendered material. Similarly, reformatting records that are in Extended Binary-Coded Decimal Interchange Code (EBCDIC) to ASCII or Unicode would cause a change in the coded representation on a storage medium, even though there is no alteration in the rendering of case files. Reformatting of electronic patent and trademark case files is likely to occur in the following circumstances:

- changing the physical carrier or storage medium
- transforming the character code from Extended Binary Code for Data Interchange (EBCDIC) to American Standard Code for Information Interchange (ASCII)
- replacing fixed-length records with variable-length records or vice versa

As a general rule, reformatting of electronic patent and trademark case files can occur independently of the operational environment within which they initially were created or maintained. To reformat electronic patent and trademark case files, the following elements are required:

- readable input storage media and operational drive
- a computer-automated reformatting program
- output storage media and appropriate drive

A computer reformatting or translation program, which typically is available as a standard utility program, performs reformatting without regard for whether the digital material consists of text files, bit map image files, structured database files, and the like. In other words, reformatting occurs without having to open the software application used to create and maintain electronic patent and trademark case files.

5.5.2 Media Renewal - Copy

The second instance of media renewal involves copying electronic records. The term "copy" in the world of paper records generally is understood to mean the duplication of the physical and

logical structures and intellectual content of an information object, as in the case of a photocopy or a carbon copy. Diplomatics definitions of a "copy" are very useful in this regard. In diplomatics, the term "copy" has three different uses. The first use, "copy in the form of the original," means that the copy has all of the same detail, including ink, paper, date, addressee, and the like, of the original, but it is produced at a later date. An "imitative copy" reproduces the content and most of the form of an original as in a photocopy. Obviously, a photocopy does not reproduce the ink and paper of the original but it does capture the substantive structure of the original. A "simple copy" is a transcription of the content of the original, but with the loss of most, if not all, of the detail of the original. A simple copy does not carry the same weight of trustworthiness as an imitative copy.

In the proposed PTO electronic patent and trademark recordkeeping application environment, what would it mean to copy an electronic case file or to copy the contents of a hard drive or a tape? The objective of copying electronic patent and trademark case files is to maintain their readability by moving them from old storage media to new storage media with the same format specifications (e.g., number of tracks and signal encoding) without any loss in structure, content, and context (i.e., an imitative copy). Essentially, in copying electronic patent and trademark case files there is no alteration or change in the underlying bit stream pattern. Copied electronic patent and trademark case files would be "binary twins," to use Ulf Andersson's phrase, because there is an exact replication of the bit stream of 1s and 0s. The requirements for copying electronic patent and trademark case files are:

- readable input storage media and operational drive
- computer duplication program
- output storage media and appropriate drive

A standard computer duplication or copy utility program performs the copy work without regard for whether the digital material is text files, bit map image files, structured database files, and the like. In other words, just as in the case of reformatting, electronic patent and trademark case files may be copied without opening the electronic recordkeeping application by which they are maintained. To summarize, copying a single electronic case file or many electronic case files

means that an exact duplication of the bit stream that gives rise to their external and internal structure and intellectual content is reproduced.

5.5.3 Transfer

The second migration methodology of ensuring that reliable electronic patent and trademark case files remain usable and accessible over their authorized retention period involves the transfer of processible electronic patent and trademark case files from a current software dependency environment to a newer one with no loss in structure, content, or context. There are three aspects of transfer that should be kept clearly in mind.

Software Dependency - the act of transferring electronic case files in one software dependency application environment, such as an electronic records management system or bit map image compression/decompression, to a new technology environment is executed within the source software application or the target software application. This requires little or no user involvement because either the source software can transfer the digital material to a neutral format that the target software recognizes and can accept, or the target software recognizes the source software and automatically executes the transfer process. which is usually called "backward compatibility." An example of the first instance would be saving WordPerfect documents in Rich Text Format (RTF), which is a platform neutral format, and then opening these documents in an Office 97 environment. An example of the second instance would be the automatic transfer of electronic documents that occurs when upgrading from Microsoft Word 6.0 to Office 97.

Change in Bit Stream - during transfer, the underlying bit stream undergoes some modification or change. This is clearly evident when decrypting electronic documents or decompressing bit map images. Although it is not as obvious, there is a change in the underlying bit stream of electronic documents because each native application software embeds non-printing ASCII characters that denote specific processing instructions, such as a hard return, end of line, and bold type, to name only a few. As a general rule, transfer differs from reformatting because it occurs either within the framework of the software application environment where electronic documents are maintained or within the target application software.

Transfer Instances - typically, transfer occurs when any of the following take place:

- Encrypted documents are decrypted or compressed bit map images are decompressed.
- The software environment is upgraded to current technology.
- A new software application environment is installed.

Because transfer occurs within a specific software dependency environment, it requires access either to the old application software or to the new application software. The technical requirements for transferring electronic patent and trademark case files are:

- A readable input storage media and operational drive
- An operational source software or target software environment
- An Established output storage media and appropriate drive

6 LONG-TERM ACCESS STRATEGY

The Patent and Trademark Office is legally obligated to ensure that electronic patent and case files are usable and accessible for forty years after their issuance and that registered electronic trademark case files are usable and accessible for an indefinite period of time until they are cancelled or expire.. In complying with this legal obligation the PTO faces two major challenges: the limited useful life of digital storage media and a constantly changing digital technology. The limited longevity of digital storage media means that over time the PTO must protect electronic patent and trademark case files from loss or deterioration and periodically move them to new storage media. The second challenge will be a constantly changing technology that will require the PTO to upgrade selected hardware components such as disk/tape drives and recordkeeping applications and to replace existing electronic recordkeeeping functions with ones that are designed for a new technology platform.

6.1 ELECTRONIC RECORDKEEPING FUNCTIONS

This section discusses six electronic recordkeeping functional areas that are required to support long-term preservation and access to usable and reliable electronic patent case file records. These six functional areas cover the requirements of being:

- Readable
- Intelligible
- Encapsulated
- Reconstructible
- Understandable
- Unaltered

6.1.1 Readable

Readable electronic patent and trademark case files mean that the bit stream comprising patent and trademark case files can be processed on the computer system or device that initially created it, the computer system or device that currently stores it, or the computer system or device that will be used to store the digital information in the future. This is a fundamental requirement because unreadable electronic patent and trademark case files serve no useful purpose. There are three areas that can affect the readability requirement

Environment. Non-readability of electronic patent and trademark case files may occur in two different ways. One is the result of exposure to a hostile storage environment. All of the storage media for digital records used in the past and present share a common vulnerability to poor environmental storage conditions. Magnetic storage media are vulnerable to fluctuations in temperature and humidity and extremes, as well as to stray magnetic fields, and they therefore require a controlled storage environment to ensure a maximum longevity. Optical digital storage media are similarly vulnerable to high temperature and high humidity.

ANSI/PIMA IT9.23-1998: Imaging Materials - Polyester Base Magnetic Tape - Storage standard recommends a stable temperature of 10° C (50° F) and a relative humidity (RH*)* of 20% as the coolest and driest conditions for preserving storage media as long as possible. Research studies conducted at the National Media Lab by John Van Bogart suggest that these same storage conditions also are appropriate for optical storage media. These are ideal conditions that may be difficult for some storage repositories to achieve. In this context, John Van Bogart recommends “equivalent" storage conditions: 20º C and 30% RH, or 15º C and 40% RH, or 10º C and 50% RH. "In humid climates, such as Florida, it would be more economical to apply storage conditions of 10º C and 50% RH. In arid climates, such as Arizona, it would be more economical to apply conditions of 20º and 30% RH." Additional protection is attained through air filtration to remove atmospheric pollutant gases (chlorine and sulfides) that are likely to be present in large urban environments.

Periodic Readability Checks. Maintaining an appropriate and stable storage environment for electronic patent and trademark case files should be supplemented by periodic checks of randomly selected storage media or files to ensure that no catastrophic media failure has

occurred or may be on the verge of occurring (36 CFR 1234.30(g)(4)). This, of course, is an early detection procedure that merely warns of potential damage. Nonetheless, a program to check annually a statistical sample of electronic patent and trademark case files for data dropouts and other potential readability problems is an essential part of ensuring the readability of electronic records.

Media Obsolescence. Attaining the predicted life expectancy of digital storage media is important, but it cannot resolve a more fundamental aspect of maintaining the readability of electronic records, which is manifested in media obsolescence. Media obsolescence occurs when the storage device (e.g. a tape or disk) is physically incompatible with the available computer hardware and therefore cannot be read. Media obsolescence seems inevitable because historically advances in storage technology have introduced changes in the way the underlying bit stream that composes digital material is physically represented. Consequently, older storage media often are incompatible with those used in the present, and surely will be incompatible with those developed in the future. There is no "final solution" to media obsolescence. The most effective way to mitigate its impact is through media renewal that consists of either periodically reformatting or copying electronic records on old media to newer media.

6.1.2 Intelligible

Any computer today can recognize a binary bit stream of 1s and 0s generated by any other computer. However, a binary bit stream is not intelligible unless it carries information about what the 1s and 0s represent that a computer can recognize and interpret for processing purposes. The intelligibility of electronic records, therefore, is a function of information about what a bit stream represents and the operating software's capacity to take appropriate action based upon this information.

Bit Map Images. The intelligibility of bit map images is a case in point. The 1s and 0s of a bit map image carry no intelligibility. However, the file header that accompanies a bit map image contains information, such as byte order and the compression algorithm used, which is absolutely essential to a computer's ability to render the image on a monitor or a printer. If these two pieces

of information are absent, most, if not all, readable electronic bit map images will not be intelligible.

Text. The 1s and 0s that comprise a textual electronic patent case file also lack intelligibility. Consider, for example, a memorandum composed in Corel WordPerfect 8.0. The document name, which is part of identification, has a three-letter extension -- WPD -- that tells a computer that this is a WordPerfect Document and the 1s and 0s are to be interpreted as ASCII code and rendered as human-readable on a monitor or a printout. A document name with a WPD extension can be read by any computer that runs WordPerfect 8.0, or by other word processing software such as Microsoft Word 6.0, which can convert a document with a WPD extension for processing. Any operating system and word processing software that recognizes 7-bit ASCII can process a document name with the TXT extension, which denotes ASCII text, without any embedded word processing instructions or code. What makes textual electronic records intelligible is located in metadata, not in the underlying bit stream of records as such.

6.1.3 Encapsulated

An encapsulated electronic patent case file means that all of the information associated with a specific electronic patent case file, such as metadata, content, and linkages to multimedia, exists as a single logical or physical entity. The distinction between a logical and a physical entity is important because techniques such as object linking and embedding (OLE) and hypertext create dynamic records that consist of pointers or linkages to files, which may exist on multiple media and servers and can only be used with specific software. Logical Encapsulation (e.g., MS WORD object linking and embedding) can be maintained and preserved in a single technology generation, but becomes increasingly difficult to achieve this across multiple technology generations.

Logical Encapsulation. Suppose, for example, that an electronic patent case file includes text, vector graphics, bit map images, and digital audio signals. In an operational environment the text portion of electronic records might exist on one file server with pointers to an image file server, a vector graphics file server, and a digital audio file server. All of this is would be transparent to users of electronic patent and trademark case files under review or as yet unpublished. When an image file server is upgraded, a message with this information is sent to

the host file server and the pointers and links are updated as necessary. Maintaining multiple file servers as part of a long-term access strategy adds several layers of complexity in configuration management. Furthermore, if all of the file servers are not regularly upgraded the risk increases of not being able to reconstruct all of the components of electronic records.

Physical Encapsulation. In contrast to logical encapsulation, physical encapsulation integrates all of the logical components of an electronic patent case file into a single wrapper or container. Physical encapsulation can help reduce the complexity of multimedia electronic records. Physical encapsulation, however, is not without its own problems. For example, the task of transferring, say, vector graphics material to upgraded or new vector graphics software could require a time intensive computer search of each electronic patent case file in the recordkeeping system to identify such material. One way of minimizing this problem would be to create a surrogate or profile about each electronic patent case file, as described below, that includes descriptive information about the different types of information representation (e.g., ASCII text, vector graphics, or bit map images) and other features that could be searched to identify components of case files that must be updated.

Standard "Open" File Format. The use of a standard open file format in encapsulation can mitigate the effects of application software or operating system dependence, which ranks high on the list of impediments to long-term access to electronic records. This form of encapsulation could employ Extensible Markup Language (XML), which is the most technology-independent solution available today to describe complex electronic patents case files and associated metadata that are essential to long-term access. A Document Type Definition (DTD) can be developed that would identify the logical structure of encapsulated electronic patent and trademark case files. For example, a DTD could be developed for electronic patent and trademark case files that delineates the hierarchical structure of patent and trademark case files, identifies multimedia material such as vector drawings and bit map images, and provides an audit trail of all actions taken to support long-term access to reliable electronic patent and trademark case files. One feature of this audit trail could include the capture of CRCs generated prior to after completion of processing (e.g., copying). All of this information would fall under a general heading of metadata to support long-term access to reliable electronic patent and trademark case files. Task

Order Deliverable 98 -03 -11Metadata Requirements For Long-term Access and Retention of Electronic Patent and Trademark Case Files proposes a metadata model for encapsulating reliable electronic patent and trademark case files.

6.1.4 Understandable

In order for electronic patent and trademark case files to be understandable they first must be intelligible to both a computer and humans. Electronic material is intelligible to humans that conveys information. One example would be a bit map record that displays the image of an airplane that humans would recognize but that a computer could recognize only with the assistance of very powerful artificial intelligence software. Another example would be ASCII text that includes discrete words, sentences, and paragraphs, which convey information that humans could recognize.

Context. In both examples, however, the larger issue is what the bit map or textual material means. The meaning of an electronic record is not determined solely by its words; meaning is also determined by the context of creation and use. As information scientist David C. Blair puts it: "The meaning of . . . a document can only be understood in its relation to other documents which, together, constitute an institutional activity like making a contract, evaluating the market for a product, [or] assessing performance." (*Language and Representation in Information Systems* (1990): pp. 210-211)

Role of Metadata. Ensuring the understandability of electronic patent and trademark case files differs sharply from ensuring the understandability of paper patent and trademark case files. Unlike paper records, which typically convey the context of creation and use in the physical characteristics of the records, the context of creation and use of electronic records are usually linked logically rather than physically. These logical linkages can include identification of both the business processes that led to the transaction and the participants in the transaction. This context of creation and use also involves relationships among other records that can be captured in a variety of ways, including a reference code in a document profile to other records dealing with the same issue, or a classification code that links records relating to the same transaction and related transactions. This captured information is in fact metadata -- electronic patent and trademark case files that lack it may not be fully understandable.

6.1.5 Reconstructible

The preservation requirement for reconstructible electronic patent and trademark case files is very important because it goes to the heart of their trustworthiness over time. Reconstructiblity means that electronic records can be rendered so that they have the "same logical and physical structure and intellectual content" as when initially stored. Thus, reconstructibility means that those internal and external features of electronic patent and trademark case files that were required for them to be complete and effective during the filing, review and examination process are also retained for the authorized retention period. In this sense, fully reconstructed renderings of electronic patent and trademark case files constitute “imitative copies,” which also helps preserve the overall authenticity and trustworthiness.

Consequences of Simple Copies. Of course, the content of electronic patent and trademark case files can be reconstructed as plain ASCII text, but archival science would then treat these as “simple copies,” which are nothing more than transcriptions whose reliability would be significantly diminished because some features that plain ASCII text cannot support would be lost. Of course, the authenticity of "simple copies" can be reestablished as new records. This would entail capturing considerable detail about the scope of the information that was lost.

6.1.6 Unaltered

Under Rule 901 of the Federal Rules of Evidence, authentic electronic patent and trademark case file records are what they purport to be -- that is, trustworthy case files whose structure, content, and context have not been altered, changed, or otherwise corrupted. Historically, protecting paper records from alteration or change was made possible through transfer of the records from the custody of their originators or creators to an independent organization or individual whose mission was to protect this material from change or alteration, such as the National Archives. This protection involved an appropriate storage environment, care and handling of the records, and controlled access to the records.

The Challenge. Unlike paper patent and trademark case files, electronic patent and trademark case files are susceptible to “alteration and corruption” in a hostile storage environment, so maintaining a stable storage environment is essential to protecting them from any loss of information. Also, electronic patent and trademark case files can be altered without leaving any

physical evidence. Furthermore, electronic patent and trademark case files may be copied and reformatted periodically to new storage media, which introduces a risk of loss or corruption either by design or by accident. The same technology that makes it relatively easy to create, use, store, and retrieve electronic patent and trademark case files also has the potential for making it possible for someone to "reinvent reality" without leaving any visible evidence of this activity. How then is it possible to ensure that electronic patent and trademark case files remain unaltered over time?

6.1.6.1 Read Only

There are several techniques to guard against this that rely upon digital information technology itself. The first employs a security feature in the electronic recordkeeping software that permits "read only" access to electronic patent and trademark case files. The net effect of this procedure is that users would only be able to download a copy of an electronic patent case file but could not save it under its existing name or a new one in the electronic record keeping system. This has the practical effect of ensuring that in electronic sense custody of electronic patent and trademark case files remains intact.

6.1.6.2 Cyclical Redundancy Checksum

The second electronic technique is called "Cyclical Redundancy Checksum" (CRC), which typically is used in telecommunications to ensure error-free transmissions. In this technique the number of bits in a packet or Frame Check Sequence (FCS) is divided (using a non-carrying division process) by a pre-defined 16-bit or 32-bit polynomial and the resulting computation is added to the FCS as a CRC. At the end of the transmission of the packet or frame, the FCS is divided by the same 16-bit or 32-bit polynomial. If the two CRCs match then there is a high probability that no error has occurred. The CRC has a high probability (99.98 per cent) of detecting all single and multiple bit errors (burst errors) that do not exceed 16 bits.

Appending a separate CRC to the frame sequences that comprise an electronic patent case file can also be used for internal processing. CRC use during media renewal (copy and reformat) and transfer can verify that no errors occurred. With the appropriate software, the resulting CRC values for a patent case file could be appended to or embedded in a record or aggregation profile as part of

a history profile or encapsulated as metadata that documents error-free processing. Of course, each time reformatting or transfer occurs, a different CRC would be computed because each one of these activities involves a change in the underlying bit stream. Subsequently, if it becomes necessary to provide a copy of one or more electronic records, a new CRC could be computed and compared to the most recent CRC in the history profile. A match would verify that no change has occurred in the underlying bit stream while in storage.

6.1.6.3

Hash Digest

The third electronic technique that can help ensure the reliability of electronic records is called one-way hashing. A one-way hash function employs an algorithm that compresses a digital object (e.g., text of a book, report, or letter, image, audio, etc.) into a fixed-length record. The two most widely used hash digest algorithms – MD5 and SHA – reduce a source "document" of any size to 128 and 160 bits, respectively. It is called a one-way hash because the hash digest itself is irreversible: it is not possible to reconstruct the original material from the hash digest. No matter how many times a digital material is "hashed," identical hash digests of the source material will be produced as long as no change has occurred. Thus, computing a second hash digest of one or more logical components of an electronic patent case file or for the complete case file itself and comparing it with the original hash digest will disclose if there was any change in each logical component or the entire case file. However, when the underlying bit stream is altered, as in the case of reformatting and transferring electronic patent and trademark case files, a new hash digest must be computed and added to a history profile.

6.2 REFORMATTING, COPYING, AND TRANSFERRING

As discussed earlier, the general concept of "migration" is addressed in this report at the more detailed levels of reformatting, copying and transferring of electronic patent and trademark case files.

One of the fundamental issues a long-term access strategy for reliable electronic patent and trademark case files must address is ensuring their processibility and transferability during the

time the are in the custody of the PTO. These two objectives can be achieved through media renewal (reformatting or copying) and transferability, which involves the "automatic" transfer of electronic case files or individual components of a case file (e.g., bit map images) from a current dependency software application to a new one.

There are four questions associated with media renewal and transfer that the implementation of a long-term access strategy for reliable electronic patent and trademark case files should take into account. They are: 1) What storage medium to use? 2) When should the storage media of electronic patent and trademark case files be renewed and when should they be transferred to a new recordkeeping system?; 3) How can the authenticity of reformatted, copied, and transferred electronic patent and trademark case files be guaranteed? and 4) How can the security of electronic patent and trademark case files in an electronic recordkeeping system be protected from intentional damage, deterioration, and catastrophic failure or natural disaster?

6.2.1 What Storage Medium to Use?

The selection of a storage medium for long-term storage of electronic patent and trademark case files is critical because this will affect the usability and accessibility of the case files and the cost effectiveness of the electronic recordkeeping application environment. Although Federal agencies have some discretion in the selection of a storage medium for long-storage of permanent records, CFR 1234.30(b)(1)-(7) identifies seven " . . . factors that shall be considered before selecting a storage medium or converting from one medium to another." The seven factors are:

1) The authorized life of the records, as determined during the scheduling process;
2) The maintenance necessary to retain the records;
3) The cost of storing and retrieving the records;
4) The records density;
5) The access time to retrieve stored records;
6) The portability of the medium (that is, selecting a medium that run on equipment offered by multiple vendors) and the ability to transfer the

information from one medium to another (such as from optical disk to magnetic tape or disk); and

7) Whether the medium meets current applicable Federal Information Processing Standards

It should be kept in mind that the PTO has a major investment in technology, particularly in storage media and the devices to read the media and may have additional experience and expertise that most likely is not reflected in 36 CFR 123.30 (b)(1) (7). Nonetheless, Factors (4), (5), and (6) are likely to be consistent with the PTO experience and merit brief discussion. Factor (4) can be interpreted as high storage density. Factor (5) can be interpreted as a high data transfer rate. Factor (6) can be interpreted as substantial market penetration through multiple vendors and conformance to interoperability requirements.

Application of these three media selection criteria would point toward selection of a magnetic storage medium that has:

- A very high storage capacity per unit of media - on the order of ten to fifty gigabytes.
- A data transfer rate in the range of 5MB or more per second.
- A substantial market presence with multiple vendor sources, both media and appropriate drives.
- A high probability of providing backward compatibility over multiple media generations.
- Incorporation of a storage management infrastructure, including copy, reformat utilities, that allow for ease of media renewal.

6.2.2 When to Reformat, Copy, and Transfer?

When is the best time to reformat or copy electronic patent and trademark case files to new media or transfer them to a new electronic recordkeeping application environment? There are three instances where reformatting is an appropriate action: (1) when the electronic case files are moved into the electronic recordkeeping application environment; (2) when the storage media or

drives used in the electronic recordkeeping application environment are upgraded; (3) when a pre-established point in the life expectancy is reached for of the storage medium or drive. The time for copying electronic patent and trademark case files is when they are ten years old, which is mandated by 36CFR 1234.30(g)(5). The point in time at which electronic case files are transferred to a new electronic recordkeeping application environment should be determined by the necessity to implement a new electronic recordkeeping application environment that is in the technology mainstream, improved cost-effectiveness that the new technology platform would bring, and the technical and financial resources available to the PTO.

6.2.3 How to Ensure Trustworthiness?

The trustworthiness of electronic patent and trademark case files could be challenged when they go through several cycles of being reformatted, copied, or transferred and the process is inadequately documented. Suppose that a body of electronic patent and trademark case files goes through three cycles of reformatting and no or limited documentation exists for how these processes were carried out. How can users be confident that these processes did not introduce changes --accidental or intentional -- in one or more case files?

Consider a specific situation in which electronic patent and trademark case files have been reformatted several times, but there is little or no documentation of the procedures followed. After the third cycle "errors are discovered. Were the errors in the patent and trademark case files when they were moved into the electronic recordkeeping application environment or did the errors occur during reformatting? If the latter, when did they occur? What was the nature and scope of the errors? Was it the result of media failure? What corrective action, if any, was taken?

These questions go to the heart of the trustworthiness of electronic patent and trademark case files over time. To ensure the long-term trustworthiness of electronic patent and trademark case files, the PTO should have a written quality control policy in place that mandates verification of the accuracy of all reformatted, copied, and transferred electronic patent and trademark case files. The procedures for implementation of this policy should include thorough and complete documentation of all the steps followed, including identification of the individual who actually executed the specific process, the date it occurred, the format of the data, generation of CRCs or

hash digests to confirm that no bits were lost or corrupted, and visual comparison of several of the "old" case files with the ones reformatted, copied, or transferred. Inaccuracies or unrecoverable errors should be identified. The block location of any unrecoverable errors should be identified. In addition, a third party should review these actions and concur that they were carried out in accordance with established procedures. This documentation should be clearly identified with links to the specific case files and treated as metadata that merits the same care as that of the case files themselves.

6.2.4 How to Protect From Alteration or Loss?

The fourth and last question to be considered is how to protect electronic patent and trademark case files from alteration or loss. Electronic storage media are vulnerable to human intrusions and catastrophic failure or natural disaster. The threat of human intrusion can be minimized by instituting access controls that permits "read only" retrieval and viewing for authorized individuals. Housing electronic storage media in a secure vault with controlled access provides another level of protection. In the event of catastrophic failure and natural disaster (such as fire or a flood) there would be little difference in the vulnerabilities of optical or magnetic media. The best security against loss of electronic patent and trademark case files through a catastrophe is keeping a second copy, called a vital records or backup copy, at a different geographical location (36 CFR 1234.30(f)). Given the cost of storage media and data transfer rate of optical vis-à-vis magnetic media, a magnetic backup is less costly. One variation on the backup security copy is to use one type of digital storage medium for primary storage and a different digital storage medium for backup storage. This can minimize the risk of loss of records if a particular storage medium suddenly and unexpectedly becomes obsolescent.

6.3 NON-PROPRIETARY STANDARDS

A fundamental premise underlying this long-term access strategy is that the implementation activities associated with it should be designed to take advantage of non-proprietary digital information technology standards that largely are products of the drive toward open systems over the last decade or so. Open systems, of course, imply standards to which individual products

conform and thereby provide relatively stable market solutions that are scalable. Non-proprietary information technology standards are especially important because they help support open systems, applications connectivity, and document portability, which in the long run may significantly enhance the prospects of long-term access to electronic patent and trademark case files. Vendors have a vested interest in retaining their customer base and therefore are likely to continue to support and develop products that are compatible with existing ones. Under this rubric, as new technologies are introduced over time, new standards will be established that link the new technologies with the old technologies. Adhering to non-proprietary information technology standards that promote interconnectivity and portability plays a key role in the long-term access strategy this study proposes. Because these standards minimize and in some instances may eliminate data exchange and computer incompatibilities, they can support interoperability and backward compatibility. Interoperability and backward compatibility are crucial to ensuring the long-term processibility and transferability of electronic patent and trademark case files.

The non-proprietary information technology standards for file formats are discussed in the report entitled *Assessment of File Formats to Support the Life Cycle Management of Patent and trademark case files* (Task Order Deliverable 98-03-08).

A cautionary note regarding information technology standards is needed, however. Standards reflect current technology and inevitably new standards will emerge to replace old ones. Therefore, continuous and careful attention must be given to the procurement of new hardware and software to ensure that they conform to non-proprietary standards that support application connectivity and document portability. The PTO should simultaneously be aware of the development of new technologies and standards that may supplant choices being made today.

7 RECOMMENDATIONS

This section proposes six recommendation areas for consideration by the PTO to support long-term access to electronic patent and trademark case files during PTO's custody. These guidelines are informed by the following considerations:

- the importance of a clear written policy statement regarding long-term access to reliable electronic patent and trademark case files
- the importance of preserving the reliability of electronic patent and trademark case files
- the need for clearly defined quality control procedures
- the actual practices for the PTO to follow in providing long-term access to reliable patent and trademark case files
- the importance of using non-proprietary standards in facilitating long-term access

7.1 GENERAL POLICY

A long-term access policy for electronic patent and trademark case files is important because it is an organizational commitment rather than an individual one. It ensures the trustworthiness of electronic patent and trademark case files as evidence, which can be achieved in part by consistent and uniform treatment and handling. Another reason for developing a formal access policy is to prepare for evidentiary questions raised in many legal proceedings: were the records kept and used in the normal course of business. This is typically established by a written policy statement.

A sound policy statement should specify the objectives of the program and the circumstances under which media renewal (reformatting, and copying) and transfer activities are initiated, and the types of compliance audits that will take place. Another issue that a policy statement should address is whether the PTO Records Management staff in conjunction with staff of the Chief Information Officer will be responsible for ensuring that all of the actions necessary to support a

long-term access strategy are carried out. A policy statement that articulates the boundaries within which specific actions may be undertaken can help establish the grounds for future users to have confidence that the electronic patent and trademark case files in fact are trustworthy

RECOMMENDATION: **Develop a PTO written policy for long-term access strategy for electronic patent and trademark case files that addresses the following:**

- **An agency commitment to comply fully with the CFR requirement to ensure that patent and trademark case files remain usable while they are in the legal custody of the PTO**
- **Protection of the integrity of electronic patent and trademark case files regardless of changes in technology**
- **Media renewal and records transferability as a means of ensuring the long-term usability of electronic patent and trademark case files**
- **Internal compliance audits for conformance with all policy and procedures**
- **Oversight of the life cycle management of electronic case files by the PTO Records Officer in coordination with CIO staff**

7.2 QUALITY CONTROL

Archival science and the Federal Rules of Evidence give the greatest weight of trustworthiness to records that are created and used in accordance with established processes and procedures. This is a form of quality control that should be carried over into the preservation of electronic patent and trademark case files. Whatever actions the PTO takes to ensure long-term access to electronic patent and trademark case files should be carried out in accordance with written PTO policies and procedures that will need to be developed. Because the objective of such procedures is to ensure the reliability of electronic patent and trademark case files, the records generated as a result of carrying out quality control procedures should be placed in an encapsulated wrapper (as metadata) that contains the associated case file. This will help ensure that documentation about quality control activities will be maintained with the same care and concern as the patent and trademark case files. Among the quality control techniques to be considered are comparisons of Cyclical Redundancy Checksums, hash digests and bit/byte comparisons which assure that no

bits or bytes have been lost during media renewal or transfer. To ensure compliance with these procedures periodic quality control audits should be conducted.

RECOMMENDATION: Develop written quality control procedures that take into account the following:

- **Documenting, and storing as metadata, any action taken to facilitate long-term access to electronic patent and trademark case files**
- **Inserting documentation records into the encapsulation wrapper that contains each electronic patent case file**
- **Conducting scheduled quality control audits**

7.3 ELECTRONIC RECORDKEEPING FUNCTION

It will be recalled that the discussion of the recordkeeping function in Section 5 took note of the fact that an AIS could serve this purpose as long as it supported patent case file preservation requirements for capture of relevant metadata that documents all actions taken to ensure the processibility and transferability of patent and trademark case files. That same discussion also suggested that at some point in the future an integrated, stand alone, electronic recordkeeping system could be a more cost effective means of supporting these preservation requirements and capture of metadata, especially considering the period of time that electronic records will be in the legal custody of the PTO.

RECOMMENDATION: Develop and implement recordkeeping functions that can be implemented within an AIS or as part of a stand alone recordkeeping system that supports:

- **Preservation of electronic patent case file integrity and authenticity**
- **Acquisition and preservation of all metadata identified in Task Order Deliverable 98-03-11.**
- **Retrieval of electronic case files based upon a data elements (properties) in profile metadata or full text**
- **Encapsulation of all components of an electronic case file**

- **Utilization of all file formats employed in the filing, review, examination, issuance, maintenance, and preservation of electronic patent and trademark case files**
- **Retrieval of electronic case files based upon a data elements ("properties") in profile metadata or full-text search**
- **Verification of audit trails of decisions and actions taken to ensure long-term access to electronic patent and trademark case files**
- **Expansion of profile metadata to include information useful for preservation purposes, such as a history use file that contains specific documentation (e.g., CRCS or hash digests) generated during the reformatting, copying, or transferring of electronic case files**

7.4 REFORMAT, COPY, AND TRANSFER

Electronic patent and trademark case files must remain unaltered and usable, despite changes in information technology. In order to address this requirement, a long-term access strategy is proposed that includes a component to ensure the processibility of electronic patent and trademark case files in a current electronic recordkeeping environment and their transferability from this dependency application environment to a new one. The specific means proposed for achieving these objectives are media renewal, which consists of reformatting and copying electronic patent and trademark case files, and transfer.

RECOMMENDATION: **Develop and implement procedures for reformatting, copying, and transferring electronic patent and trademark case files that include the following:**

- **Reformat electronic case files when new storage devices or media are utilized**
- **Copy electronic patent and trademark case files at the time they are moved into the electronic recordkeeping application environment**
- **Copy electronic patent and trademark case files every ten years (36 CFR 123430[g)(5) if magnetic tape is the storage medium**
- **Copy electronic patent and trademark case files when the annual readability sample discloses ten or more temporary or read "errors" on a specific storage medium**

- **Transfer electronic patent and trademark case files to a technology platform when the current software is upgraded or a new electronic records management application environment is installed**
- **Ensure the reliability and integrity of reformatted, copied, and transferred electronic patent and trademark case files by employing a strict quality control procedure that may include bit/byte comparisons and comparisons of hash digests and Cyclical Redundancy Checksums (CRC)**
- **Document fully all actions taken when reformatting, copying and transferring electronic patent and trademark case files and include this information in the metadata associated with each case file**
- **Create two copies of electronic patent and trademark case files at the time of reformatting, copying, or transfer and store one copy at an off-site location.**

7.5 PHYSICAL SECURITY

Protecting the physical security of electronic patent and trademark case files from alteration, change, or loss should be a critical component of the mission of a PTO storage repository in which electronic patent and trademark case files are stored. Achieving this objective involves several separate but related tasks.

One aspect of physical security is using a storage facility with controlled access. This means that access to the facility is limited to authorized staff and a record (sign in/out log) is kept of the date, time, and identity of each person who enters it. A senior manager should review this log periodically to verify staff compliance, and the log itself should be retained as evidence of the storage repository's compliance with is own policies.

The storage repository should be located where the threat of a natural disaster such as a flood, fire, or earthquake is minimal. The storage repository should include a fire detection alarm system as well as a water sprinkler fire suppressant system. Individual storage media should be sealed in polyethylene bags for protection from water damage when the water sprinkler system is activated. A full-scale disaster recovery plan should be in place that can be quickly implemented.

If a magnetic storage medium is selected for long-term storage of electronic patent and trademark case files, its vulnerability to magnetic fields must be taken into account. The actual storage repository should be located away from large, heavy duty electric motors (e.g., high-capacity air conditioning systems), generators, transformers, and electrical power lines carrying high current loads because unless they are magnetically shielded they can emit a strong magnetic force. Signs should be posted warning individuals entering the storage repository that all magnetic devices are prohibited and failure to comply will result in severe penalties. Even though a distance of 76 mm (3.0 inches) is likely to protect magnetic media from erasure emitted by most electric devices that may be required in a storage repository, it would be prudent to keep electric motors and similar devices at least two feet from magnetic storage media. There is no danger of erasure when magnetic storage media themselves are stored close together.

There is no guarantee that any one or all of these proposed approaches will provide full protection of electronic patent and trademark case files from alteration, change, or loss. Consequently, the PTO storage repository should have a recovery alternative in place that can provide protection in the most dire of circumstances. A "fail safe" recovery alternative is to maintain at a second site, preferably in another geographic area, a backup "mirror" copy of all electronic patent and trademark case files in the storage repository. Of course, some electronic patent and trademark case files might be deemed so important to warrant having another "mirror" copy at a third site. This backup "mirror" copy could be used to restore any electronic patent and trademark case files that may have been lost or corrupted. This restoration would have to be fully documented and included in the actions that have been taken to ensure the authenticity of the patent and trademark case files.

RECOMMENDATION: **Develop written physical security protection procedures that take into account the following:**

- **Maintaining a controlled access storage facility and monitor compliance with access control procedures**
- **Locating the storage facility in an area where the threat of a natural disaster is minimal**

- **Creating a disaster recovery plan**
- **Adhering to the recommendations of *NBS Special Publication 500-101 Care and Handling of Computer Magnetic Storage Media* (pp. 37 - 52) regarding exposure to magnetic fields if a magnetic storage medium is used**
- **Maintaining a second storage facility where "back-up" copies of all electronic patent and trademark case files are stored**

7.6 ENVIRONMENTAL CONTROL AND MONITORING PROGRAM

A controlled environment and monitoring program is an essential part of protecting the readability of electronic patent and trademark case files. Magnetic media are inherently unstable, so their longevity and readability are at risk in a hostile storage environment where the temperature and humidity are substantially elevated. Although, optical media in general are somewhat more stable than magnetic media, nonetheless they too are at risk when exposed to high relative temperature and humidity. One of the primary reasons for this vulnerability of magnetic and optical storage media is that the binder that holds magnetic particles to a substrate (as in magnetic recording fabrication) and optical media recording material to a substrate are susceptible to 'binder hydrolysis" when exposed to high relative humidity.

National Media Lab Research. Research conducted by the National Media Laboratory on predicted life expectancy of magnetic and optical storage media confirms the negative impact of elevated temperatures in excess of 26 º C (74° F) and relative humidity of 70 percent. According to John Van Bogart, the Principal Investigator for Media Stability Studies at the National Media Laboratory who conducted these studies, 10 º C (50° F) and 20 percent relative humidity provide the ideal storage environment for electronic records to be retained for long periods of time. Van Bogart acknowledges that this ideal storage environment may not be feasible for many storage repositories, so he advises that 15 º C (59±5° F) and 40% maximum relative humidity " are safe practical storage conditions." It must be underscored that these are stable levels and any variation above the recommended temperatures and humidity ranges is likely to significantly reduce media life expectancy.

Cleanliness. Another aspect of a program for the care and handling of electronic patent and trademark case files in long-term storage involves cleanliness. Airborne dust, food, and smoke particles deposited on the surface of the storage media can cause a loss in the signal during playback. Typically, these are temporary errors that can be corrected by cleaning the surface of the storage media. The important point, however, is to avoid temporary "read" problems or "data dropout" by using air filters to remove dust particles and by prohibiting food, beverages, or smoking in the storage facility. Another source of "data dropouts" is any oil deposit from fingers, which attracts more dust, so lint-free cotton gloves should be worn when it is necessary to touch the tape surface.

Periodic Inspection. The last component of a controlled environment and monitoring programs for electronic patent and trademark case files is a periodic inspection of selected storage media to ensure that no catastrophic loss has occurred or may be impending, by selecting a statistical sample of media to be read. The number of items in the sample depends upon the total number of storage media held in the storage repository. 36 CFR 1234.30 (4) specifies the following guidelines for selecting media to be included in the annual statistical sample.

- Select all storage media annually if they are fewer than fifty
- Select a 20% random sample of the storage media when the total number of storage media ranges between fifty and 1800
- Select a random sample of 384 storage media when the total number of storage media is 1801 or greater

<u>RECOMMENDATION</u>: Ensure the continued readability of electronic patent and trademark case files by establishing a program that provides (1) a stable storage environment, and (2) good care and handling procedures. Such a program should include the following:

- **Maintaining a stable storage environment in which the temperature and relatively humidity (RH) are 15º C (59° F±5) and 40% (±5)**
- **Installing a filter system to remove air dust particles and gas pollutants**
- **Prohibiting the consumption of food and beverages and smoking in the storage facility**

- **Implementing a program based upon 36 CFR 1234.309(g)(4) to read annually a statistical sample of the storage media to identify real or impending loss of information**