Produced for:
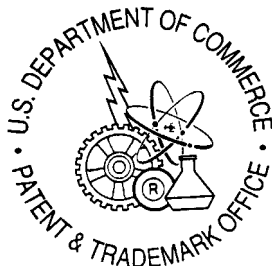
U.S. DEPARTMENT OF COMMERCE · PATENT & TRADEMARK OFFICE

# METADATA REQUIREMENTS

## FOR LONG-TERM ACCESS AND RETENTION OF

## ELECTRONIC PATENT AND TRADEMARK CASE FILES

February 26, 1999

**FINAL**

# Table of Contents

# 1  INTRODUCTION

Between now and the year 2003, the Patent and Trademark Office (PTO) is undertaking an electronic patent and trademark application filing, processing and maintenance program that is part of the "Reinvention Goals for 2000" which establishes the PTO's vision for the 21st Century - - " to lead the world in providing customer-valued intellectual property rights that spark innovation, create consumer confidence and promote creativity."

The PTO recognizes that many of the policies and procedures designed and implemented to manage paper-based case files (and their contents) do not translate easily into the realm of electronic case files because electronic records only exist as physical entities when they are rendered on a monitor or printer.  Consequently, there is a need to identify the high level issues related to the management of electronic patent and trademark records over the full life cycle, and to develop recommendations for resolution.

One area where the need is particularly important relates to the identification and acquisition of descriptive information that supports accessibility to electronic patent and trademark case files, and provides for protecting the integrity of the case files over the long term.   Another need is the identification and capture of information that documents what actions, if any, the PTO has taken to insure the usability and authenticity of electronic patent and trademark case files during their authorized disposition period.  This descriptive information about the electronic case files and the records they contain is referred to in this report as "metadata".

Traditionally, the term "metadata" has been widely used to characterize the descriptive information that will support search and retrieval of both paper and electronic material.  Over the last three or four years the use of the term metadata has expanded to include additional information that must be acquired and retained in order to effectively manage electronic records over long period of time, including permanently.

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.

The PTO will have to manage a growing volume of electronic patent and trademark case files on an on-going basis far into the future.  Consequently, this study is being undertaken to identify the metadata requirements for effectively managing patent and trademark case files for long term access and retention.

The study consists of seven sections.  Section 1 contains introductory and background information about the document.  Section 2 defines the purpose and scope of the deliverable.  Section 3 is a management summary.  Section 4 reviews the history of metadata and the findings of several metadata research projects and reports.  Section 5 discusses several metadata implications for the life cycle management of electronic patent case files.  A proposed metadata model for the life cycle management of electron patent case files is the focus of Section 6.  Section 7, which proposes several recommendations for the PTO to consider, rounds out the study.

# 2  PURPOSE AND SCOPE

The purpose of this report is to identify metadata requirements that support the long term management and use of electronic patent and trademark case files.

This document identifies specific metadata required to support the long-term access and migration strategy (as defined in Task Order 98-03, Deliverable 98-03-10) which includes assuring usable and authentic electronic patent and trademark case files for the authorized disposition period. The identification of metadata requirements is based upon a review of relevant sections of the CFR, the metadata specified in DOD 5015.2, and the recommendations of three major metadata projects and studies.  The metadata requirements also take into account broader metadata implications such as:

- Core metadata that is absolutely essential to ensure long term accessibility and protect the integrity of electronic patent and trademark case files.

- The integration ("encapsulation") of all the metadata with the  electronic case file or file wrapper.

The metadata requirements and definitions in this study may be refined where it is determined, in reviews with the PTO, that modifications and additions need to be made to better fit with the PTO's existing and planned electronic file wrapper definitions and processes, and with the information technology infrastructure.

# 3   MANAGEMENT SUMMARY

Organized metadata, that is data about data, has a long history. Metadata in the form of bibliographic data recorded on library catalogs that facilitate access and retrieval of books and other published documentary material have been in place for more than a hundred years. The emergence of powerful computer technologies over the last three decades has drastically altered the requirements for and the capabilities for capturing, storing, retrieving, and using metadata. The growth of the Internet with its focus on accessibility to digital material stored on different technology platforms has added a new level of complexity.

This study examines the metadata requirements identified in appropriate sections of the Code of Federal Regulations, Department of Defense 5015.2, the "Requirements for Business Acceptable Communications" in the University of Pittsburgh study entitled *Functional Requirements for Electronic Records*, and a metadata reference model in a study conducted by the Society of Motion Picture and Television Engineers. These studies share in common the understanding that the level of assurance about the usability and trustworthiness of paper records is not easily extended to electronic records, which exist as logical entities and must be interpreted by software and then rendered into an intellectual form that humans can understand and use. Metadata, they argue, forms the nexus between electronic records as logical entities (the logical structure of the content) and records as a physical structure (how they look). Metadata also provides the underlying information required to access and retain electronic case files for the scheduled retention period. Metadata, therefore, is a crucial component in the life cycle management of electronic records.

## 3.1   IMPLICATIONS

There are four critical metadata implications for the long-term management- of electronic patent and trademark case files.

**Usability**.  The first issue involves the continued usability of electronic patent and trademark case files, which is mandated by 36 CFR §1234.24 (b) (1) (ii).  The current thinking within the records management and archives communities is that information at both the case file level and the discrete record level that enables them to be processed and rendered as intelligible and understandable to humans must be captured and maintained along with the case files.  Sometimes referred to as "history metadata," it documents the file format, storage media, software (operating systems and applications), and other procedures used in the life cycle management of electronic patent and trademark case files.

**Integrity.**  Protecting the integrity and thereby the authenticity of electronic patent and trademark case files is a key legal (and credibility) requirement for the PTO.  Therefore, metadata that can help ensure the integrity of electronic patent and trademark case files and the individual records comprising them is a fundamental requirement.  The capacity of metadata to document that there is no loss in the structure, content, or context of electronic records when they are exposed to a potential risk of loss or alteration during media renewal or transfer to a new technology platform can help satisfy this requirement.  This is particularly noteworthy because such metadata documentation constitutes an "audit trail" that can be used to identify and reconstruct the result of all activities that have a potential risk exposure.

**Wrappers.**  The PTO file wrapper used with paper patent and trademark case files serves several important purposes.  First, it distinguishes a case file as a physical entity.  Second, while a case file is pending the wrapper can aid in protecting the confidentiality of the contents from unauthorized access or viewing.  Third, it allow the contents of a case file to be arranged according to a prescribed scheme that facilitates quick access to specific records.  The file wrapper, therefore, makes paper patent and trademark case files "self-referential" in the sense that no additional explanatory or linkage information is required.

Electronic patent and trademark case files exist as logical entities, not physical entities, so they lack the "self-referential functionality" of paper patent and trademark case files.  Consequently, the only way known at this time to support this "self-referential functionality" of electronic patent and trademark case files is through the use of metadata.  Building upon the metadata reference model of the Society of Motion Picture and Television Engineers (SMPTE), this study

identifies metadata requirements that will support this "self-referential functionality" in the life cycle management of electronic patent and trademark case files.

**Logical vs. Physical Encapsulation.**  An electronic case file can be used to encapsulate or integrate all of the records, including those in different formats, that comprise a case file into a single digital object.  Encapsulation of a patent or trademark case file can be either as a logical or physical entity.

In logical encapsulation all of the information comprising the metadata and content would be linked but potentially stored on different file servers and different volumes of media.  For example, bit map records might be stored on an image file server, vector drawings, on a vector file server, text files on a text file server and metadata on a database server.  Although logical encapsulation minimizes centralized storage requirements, it could involve very complex configuration management to minimize the loss or corruption of points that decouple elements of a record or a case file.

Physical encapsulation involves writing all of the metadata and content information within a "case file wrapper" on a single volume of media where the case file is stored as a single physical entity.  Physical encapsulation minimizes the risk of pointers and linkages being corrupted during media renewal or transfer that could result in the inability to reconstruct fully the entire case file at some subsequent point in time.  At the same time the potential requirement of considerable incremental processing time required to locate and convert selected records within case files to a new file format must be considered.

The decision about logical or physical encapsulation of electronic patent and trademark case files should be made within the context of the functionality of the planned AIS and the degree to which supporting the maintenance phase of the case files might degrade the performance of AIS or otherwise diminish cost-effectiveness.

## 3.2   CORE METADATA REQUIREMENTS

The core of this study is a section called Metadata Requirements that proposes six categories of metadata to support the long-term access and retention of electronic patent and trademark case files.  The six categories are:

- **Case File Profile**: Stores information about each "encapsulated" case file  to support long-term retention and management

- **Case File Record Profile**: Stores information about each record in a case file

- **Use History Profile**: Stores tracking information about all of the activities and actions associated with each  case files after issuance or abandonment

- **Reformat Profile**: Stores detailed information about a case file before and after reformatting to ensure its trustworthiness

- **Copy Profile**: Stores detailed information about a case file before and after copying to ensure its trustworthiness

- **Transfer Profile**: Stores detailed information about a case file before and after transfer to ensure its trustworthiness

The Reformat, Copy, and Transfer profiles are based upon the analysis and definitions established in the report produced for Deliverable 98-3-10, *Long Term Access and Migration Strategy for the Life Cycle Management of Electronic Patent and Trademark Case Files*, dated 2/26/99.  The specific metadata elements that comprise each of the profiles are reviewed in some detail in Section 6 of this report.

The descriptions of the six metadata profiles do not stipulate how each element should be captured.  Nevertheless, it is assumed that some data elements, such as those for the Case File Profile and Case Record Profile, can be automatically acquired from the updated PALM and TRAM replacement databases. Some metadata elements such as access dates are assumed to be supplied automatically by the AIS, or as a last resort by manual data entry.  Other  data elements from these two profiles could automatically "populate" the Use History, Reformat, Copy, and Transfer Profiles when they are first opened.

The metadata profiles are designed so that there are clear linkages of individual metadata elements within and between profiles. For example, the Use History Profile contains a metadata element called "Access Dates" that is a sequential date list of every access to a case file. These "Access Dates" could be linked to the dates when a specific record was accessed or to the dates when an entire case file or individual records were reformatted, copied, or transferred.

These six categories of patent and trademark case file metadata profiles should be integrated into the design and planning for PALM and TRAM database replacements as well as other electronic patent and trademark AISs and databases. The elements and properties of these metadata profiles may require more explicit definition and refinement during this process. However, the basic objective of these metadata profiles to provide documentation for an "audit trail" that can be used to identify and reconstruct the result of all activities that have a potential risk exposure must remain unchanged.

## 3.3 RECOMMENDATIONS

This study proposes three recommendations that can help assure that patent and trademark case files will be both usable and trustworthy no matter what technology changes may occur. In addition to the metadata requirements identified in the relevant sections of the CFR, DOD 5015.2, and the findings of three major metadata research projects and studies, these recommendations take into account broader metadata implications for the PTO that include:

- **Technology platform neutrality**, which means that the metadata and case file content must be capable of being read by any computer without regard for the computer or application software used originally to create them.

- **Compatibility**, which means that the electronic case file "wrapper" formats must be compatible with existing file formats and established working practices.

- **Extensibility**, which means that the electronic case file "wrapper" format along with the metadata and record content should have a longevity of at least one decade or more and

must be extensible in the sense that the "wrapper" format must be capable of accommodating new file formats and metadata types.

The three recommendation are:

## 1. Review, Refine, Integrate, and Prototype

Review, refine, and integrate the proposed metadata profiles into the overall metadata and electronic file wrapper plans of the PTO and conduct a prototype evaluation. The prototype would include conducting tests of the proposed Case File and Case File Record profiles within the context of overall patent and trademark metadata and electronic file wrapper management and determining the effectiveness of the metadata profiles for Use History, Reformat, Copy, and Transfer.

The goal of the review, refinement process and the integration and prototype tests is to determine the level of compatibility with the current definitions and plans for an electronic file wrappers and to determine the degree of compatibility with the PALM and TRAM upgrades or replacements.

Another goal of the prototype process should be to test the ability of the metadata requirements to support the usability and preserve the integrity of electronic patent and trademark case files and all records associated with them.

Over the retention periods for patent and trademark case files, the volume of metadata generated under each of the six Metadata Profile categories could be quite substantial and, if this were the situation, it could create a significant overhead cost to the operational program. An assessment of the storage and processing costs that are likely to be associated with the capture and maintenance of the metadata profiles metadata should be part of this prototype evaluation.

## 2. Develop A Common Metadata and File Wrapper Foundation

The discussion of an electronic patent and trademark case file "wrapper" in this study is shaped largely by what is happening in areas that are somewhat removed from current PTO day to day

operations.  However, over time the PTO will have to take these developments into account as it plans new information technology programs.

Three PTO information technology programs, the PALM and TRAM upgrades or replacements and TEAM, are in the process of defining and developing the foundation for certain metadata and electronic file wrapper management .  It is important that the metadata profiles and properties, and the concept of electronic case file management defined in this report are taken into account in both the patent and trademark automation programs.  Based on the similarities of requirements for long term management of patent and trademark electronic case files, consideration should be given to developing a common metadata and electronic file wrapper foundation that could serve the needs of both the patent application and trademark electronic work places.

## 3.   Conduct a Cost Benefit Analysis of Logical and  Physical "Encapsulation"

This study has identified several potential risks and benefits for employing either a physical or logical encapsulation of the metadata profiles into the electronic case file or file wrapper.  The identification of risk and benefits, however, is at an abstract level and does not take into account the specific information technology infrastructure of the PTO.  For example, it may be that the existing information technology now in place, along with planned upgrades, is such that only logical case file encapsulation makes sense either from a financial perspective or based on the information technology infrastructure.

The decision on the implementation of a logical or physical case file wrapper must be evaluated in a more in-depth study that is based on the costs and benefits as they relate to: a) existing and planned technology for PTO server infrastructure and storage management strategy, b) planning and implementation of the PALM and TRAM upgrades or replacements, c) the definition and implementation direction and status of the "electronic file wrapper", and d) the definitions and recommendations in this study for metadata profiles and case file encapsulation methodology for supporting long term management of electronic patent and trademark case file records.

# 4  METADATA – BACKGROUND

Metadata, which traditionally has been understood to mean "data about data," is a term that describes a function that has undergone some refinement and modification over time.  In the library world the attributes of books (e.g., author, title, and publisher) have been associated with descriptive cataloging, the purpose of which was to facilitate access and retrieval of books and other published material.  Of course, this form of descriptive cataloging was based upon the standard library catalog card used since the late 19th century in North America to enhance access to specific library material.  In the 1960s and 1970s as data libraries and data archives came into existence, one of their key functions was to assemble the documentation and codebooks that explained the contents of "machine-readable data files."  Codebooks were essential because they frequently contained background material on the methodology of data collection and explained arbitrarily assigned values to demographic attributes (e.g., 1 for male, 2 for female) in data files that were not intuitively understandable.

The importance of this kind of documentation increasingly became more crucial in the collection and analysis of spatial science data collected by satellites and space exploration probes.  The specific values collected by scientific instruments could not be fully understood or relied upon without having additional information about the time of collection, the orientation of the instruments, their history of perceived accuracy, and the algorithms used to transform instrument readings into values that could be used in scientific analysis.  Thus, in the space science community this "metadata" became a necessary vehicle for establishing the understandability and reliability of spatial science data.

The widespread use of Internet services over the last six years or so has greatly heightened the importance of metadata as descriptive tool that facilitates the discovery of and retrieval of digital resources.  Digital objects that may be located and retrieved through Internet services may be in a variety of formats, some of which may be unreadable in a specific technology environment without translation.  In other instances, there may be restrictions on access, such as payment of a

fee, or the desired resource is quite large and would require a very long time to access or download.  Metadata to support Internet services, therefore refers to information about digital resources that can facilitate their location and retrieval and also can support interoperability among and between heterogeneous computer platforms.

Sometimes called a document profile, these characteristics may include information about the creator or author of the resource; why, when, and how the resource came into existence; data collection methods, if any; accuracy of the source data; source granularity, processing and use history; and the quality and extent or scope of the resource.  Users of Office 97, Corel WordPerfect 8.0, and VISIO 5, among others, who create a document summary in fact are creating metadata.  For example, the default document properties screen in Office 97 includes title, subject, author, manager, keyword, and content.  There are twenty-six optional fields, including date created, date modified, date last accessed, and statistics (number of words, sentences, and paragraphs) that can be added to the document properties.  These four particular properties, it should be noted, require no user action because they are automatically extracted from the document or a use history file.  The document properties can be used to identify the circumstances of creation, to facilitate retrieval by any of the discrete document properties because they serve as index or sort terms, and to support limited audit trail activities about date of creation and date of most recent modification and access.

In the context of the PTO, metadata can be defined as information that describes characteristics of an electronic resource case file and its contents which can be used to retrieve one or more case files based upon selected characteristics, to identify and document associated information such as audit trail data related to migration or pointers to physical storage media.  The PTO already has databases that act as "metadata" in that PALM and TRAM provide a form of bibliographic and status data related to the paper-based patent case files.

## 4.1 REGULATIONS AND GUIDELINES

### 4.1.1 Code of Federal Regulations

Although the parts of the Code of Federal Regulations (CFR) that bear on the management of electronic records do not explicitly discuss metadata, the overall purpose of metadata in electronic records management is clearly articulated. 36 CFR Ch. XII Subpart C- Standards for the Creation, Use, Preservation, and Disposition of Electronic records includes the following requirements that can should be construed as metadata requirements:

> "Agencies shall maintain adequate documentation and up-to-date technical documentation for each electronic information system that produces, uses, or stores data files. Minimum documentation required is a narrative description of the system: physical and technical characteristics of the records, including a record layout that describes each field, including its name, size, starting or relative position, and a description of the form of the data (such as alphabetic, zoned decimal, packed decimal or numeric), or a data dictionary or the equivalent information associated with the data base management system including a description of the relationships between data elements in the data base." [36 CFR §1234.20(b)]

> ""Provide a method for all authorized users of the system to retrieve desired documents, such as an indexing or text search system." [36 CFR §1234.22 (a)(2)]

> "Appropriate identifying information for each document maintained on electronic media may include: office of origin, file code, key words for retrieval, addressee (if any), signator, author, date, authorized disposition (coded or otherwise), and security classification (if applicable). Agencies shall ensure that records maintained in such systems can be correlated with related records on paper, microfilm, or other media." [6 CFR §1234.22(b)].

> "Some transmission data (names of senders and addressee(s) and date the message was sent) must be preserved for each electronic mail record in order for the context of the message to be understood." [36 CFR §1234.24(1)]

> "Agencies that use an electronic mail system that identifies users by codes or nicknames or identifies addresses only by the name of a distribution list shall ...retain names on directories or distribution lists to ensure identification of the sender and addressee(s) ..." [36 CFR §1234.24(3)]

> "Agencies that use an electronic mail system that allows users to request acknowledgements or receipts showing that a message reached the mailbox of inbox of the addressee, or that an addressee opened the message shall issue instructions to e-mail users specifying when to request such receipts or acknowledgements for recordkeeping purposes and how to preserve them" [36 CFR §1234.24(3)]

> "....it is essential that agencies distinguish between records and nonrecord materials [36 CFR §1222.34(a)]

The CFR metadata requirements for electronic records management can be summarized as follows:

- Record or nonrecord status

- Key words and indexing

- Office of origin

- File code

- Date

- Addressee

- Signator

- Author

- Authorized disposition

- Security classification

- Transmission data for e-mail

- Identification of aliases and code names of addressees of e-mail

- Acknowledgement of receipt of e-mail messages

## 4.1.2  DOD 5015.2

Unlike the CFR citations above, the U.S. Department of Defense Directive 5015.2, Design Criteria for Electronic Records Management Software explicitly defines metadata in two closely related ways.  The first definition states that metadata is "data describing stored data: that is, data describing the structure, data elements, interrelationships, and other characteristics of electronic records."

The second definition links metadata to a record profile, which is defined as "information" (metadata) about a record that is used by the Records Management Application (RMA) to *file and retrieve* the record.

It includes information fields such as:

- To

- From

- Date

- Subject

- Unique record identifier

- Document type

- Format location

- Record number version number

- File category

- Originating organization

Other metadata requirements for the management of electronic records noted in DOD 5015.2 are:

- Media type

- Transmission and receipt data of email, including other recipients

- Vital record indicator

- Record category name and code

- Author or originator

- Date of record receipt

- File plan (unique ID, title, description, and disposition authority)

- Disposition instruction code

- Disposition cutoff date

- Disposition action date

- Disposition action code

## 4.2 STUDIES

### 4.2.1 Dublin Core

In March of 1995 the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) organized a conference held in Dublin, Ohio to identify and promote a set of commonly used and understood descriptors of electronic resources that would describe basic information about them and support the sharing of this information across heterogeneous technology platforms. Conference participants identified thirteen attributes of electronic resources that comprised the minimum descriptive information (hence "core elements") needed to achieve sharing across Internet services. In the ensuing three years there have been four other "Dublin Core" conferences that have led to an expansion of the core elements to fifteen. The fifteen core elements along with a brief description of each is given below. (See http://www.ub2.lu.se/tk/metadata/DC10cats.html)

1. **Title**: The name given to the resource by the creator or published

2. **Author or Creator**: The person(s) or organizations responsible for the intellectual content of the resource

3. **Subject**: The topic of the resource or keywords and phrases that describe its intellectual content

4. **Description**: A textual description of the intellectual content of the resources, including an abstract

5. **Publisher**: The entity responsible for making the resource available in its present form

6. **Contributors**: Identification of the person(s) or organization(s) in addition to the Creator who have made significant intellectual contributions to the resource

7. **Date**: The date the resource was made available in its present form

8. **Resource Type**: The category of the resource such as a home page, novel, working paper, technical report, and the like

9. **Format**: The data representation(s) of the resource, such as ASCII, RTF, JPEG, and the like

10. **Identifier**: A string or number used to uniquely identify the resources (e.g., ISBN)

11. **Source:** The source, either paper or electronic, from which the resource is derived

12. **Language**: Language(s) in which the intellectual content of the resource is conveyed

13. **Relation**: Expresses formal relations among resources that have formal relations to other resources

14. **Coverage**: Geographic and temporal duration characteristics of the resource

15. **Rights**:      Associate terms and conditions of use, including copyright statements with a resource

Work on refining the Dublin Core continues but within a larger context of the World Wide Web Consortium (W3C), especially with regard to the Extensible Markup Language (XML) standard and its pending extensions.

## 4.2.2   University of Pittsburgh

Although considerable attention has been directed toward metadata resources for sharing information on the Internet, metadata is not limited to a specific discipline.  In fact, the chief architect of the initial Dublin Core, Stuart Weibel of OCLC, appears to have been inspired in part by the Reference Model for Acceptable Business Communications that was produced as part of the University of Pittsburgh Research Project on Electronic Records.   (Available electronically at http://www.sis.pitt.edu/~nhprc/meta96.htlm and http://www.sis.pitt. edu/~nhprc/BAC/articl.html.)

A premise underlying the development of this business model was the need for standards of data interchange that would help ensure the integrity and trustworthiness of business communications as evidence.  Consequently, a fundamental aspect of this model is its articulation of metadata requirements that result in the capture and storage of information that can help assure the admissibility of business communications as records and their longevity as determined by retention instructions.  These metadata requirements are derived from twenty properties of records as evidence that the research identified in law, regulation, and best practice.  These metadata requirements in turn are organized into a logical sequence that software interprets in order to execute the activities necessary for records to be processed and understood.   The

underlying model is one in which records are considered objects and their associated metadata are considered objects with "class" features that are inherited from the records object.

The six layers are: Handle, Terms & Conditions, Structural, Contextual, Content, and Use History. The metadata in the Handle Layer identifies the data as a record, assigns values denoting the provenance of the record, and provides content descriptor terms for retrieval. The Terms & Conditions Layer contains metadata that controls access to and the use and disposition of a record.

The third layer - Structural - is the richest and most complex of the six layers largely because it consists of metadata about the record structure that is essential to the preservation of its authenticity over time and for the "migration" of the record to new software and hardware environments as necessary. One of the sub-components in this layer addresses the data representation used, such as ASCII or EBCDIC for character data and compression and encryption schemes. Another sub-component of this layer consists of metadata that defines the structure of the record, including authority files, version relationships, and dynamic relationships.

The Contextual Layer contains metadata that identifies the individual, system, or instrument that is responsible for generating the record and provides information that supports its use as evidence of a transaction.

The fifth layer, called Content, contains the information that comprises the transaction.

The Use History Layer, which is the sixth layer, contains metadata that documents use of the record after its creation or receipt. Among the possible uses suggested that metadata should document are the type of use, when the record was used, and by whom. Although it is not explicitly stated, the Use History Metadata could include documentation of actions taken to ensure long-term access, including reformatting and copying electronic patent case files to new media or transferring the case files to a new recordkeeping software environment.

A key feature of the Pittsburgh metadata model is that metadata is defined in such a way so that electronic records (objects) can be communicated across software and hardware layers with the

intend that records are "self-documenting, self-authenticating, self-redacting, self-migrating, and self-disposing."

It is not within the scope of this review of the reference model to replicate all aspects of it but rather to focus on salient features that could be useful to the PTO.  They include:

- Metadata that would ensure the unambiguous identity of records and any conditions attached to their use

- Metadata that can be used to retrieve records based upon specified characteristics

- Metadata with the capacity to disclose when records were created, why they were created, and how and when they were used in an operational environment, including specific hardware and software specifications.

- Metadata that notes the retention period for each record, the disposition authority, and what action (destruction), if any, was taken and by whom.

- Metadata that contains information about how the digital signals comprising the records are encoded (e.g.,  ASCII or bit map images), the non-proprietary file formats used (SGML, XML, CGM, and the like), and any compression algorithms used

- Metadata that provides information about all actions taken to ensure the continued accessability and usability of records, such as cyclical redundancy checks (CRCs) and hash digests that may be used to detect any changes in the underlying bit stream of records when they are "migrated" to new media or a  new technology environment

The proposed PTO Metadata Reference Model that is discussed in section six incorporates a number of the features of the University of Pittsburgh Reference Model for Acceptable Business Communications.

### 4.2.3   Harmonized Standards for the Exchange of Television Program Material as Bit Streams

In 1996 the European Broadcasting Union (EUB) and the Society of Motion Picture and Television Engineers (SMPTE) organized a joint Task Force for the Harmonization of Standards for the Exchange of Programme Material as Bitstreams.  The impetus for this effort was the rapid convergence of television, and computer technologies and a growing consumer demand for

programming that would require a leveraging of technologies to gain improved efficiency and economies of scale in the dissemination of this programming.  Dealing with this technology convergence and consumer demand would entail a major investment in technology but its widespread availability probably would lead to a proliferation of distribution channels and fragmentation of audiences, which inevitably would mean that smaller budgets would be available for production.

The EUB and SMPTE believed that the most effective way of countering this trend was to identify new and more efficient production tools that could leverage technology strengths through harmonization of standards and practices.  Consequently, the EUB and SMPTE established a Task Force for the Harmonization of Standards described earlier.  The final report of this task force was issued in July 1998 (available electronically as a PDF file on http://www. smpte.org/engr/ftrpt2w6.pdf).  Initially, a report dealing with television programme materials as bitstreams might appear irrelevant for a study dealing with metadata requirements to support the long-term retention of electronic patent and trademark case files.  However, Chapter 4 of the report entitled 'Wrappers and Metadata" delineates a high level conceptual model of metadata and wrappers for use in the transmission of television programs as bit streams parts of which can be extrapolated to metadata requirements that support retention of usable electronic patent records in case files.

4.2.3.1  **Wrappers and Content**.  According to the task force report " Wrappers are intended for use in linking physical media together, for streaming of programme material across interconnects [broadcast], and to store program material in file systems and on file servers." (p. 62).  Two of the aspects of Wrappers merit special attention.  First, programme material and any other related material are defined as Content, which in turn is subdivided into two categories.  One category is Essence, which consists of the actual program signals.  The second category is Metadata, which describes the Essence and any other information that facilitates accurate interpretation of Essence.  More to the point, Metadata is any content information other than Essence that has a contextual relationship to the associated Essence.  Second, when Content is assembled and placed on to a tape of disk for later access it is kept in a Storage Wrapper.  A Wrapper not only

contains Content but it also defines and describes the structure of the Content.
According to the Wrapper schema, various types of Essence may be defined, such as
video, audio, graphics, text, still images, to name a few.  This Wrapper Model is
delineated in the following schematic.

## SCHEMATIC VIEW OF WRAPPERS
## IN FILE STORAGE

**WRAPPER**

**ESSENCE**

**METADATA**

Wrapper

**ESSENCE**

**METADATA**

**MEDIA STORAGE**

**FILE STORAGE**
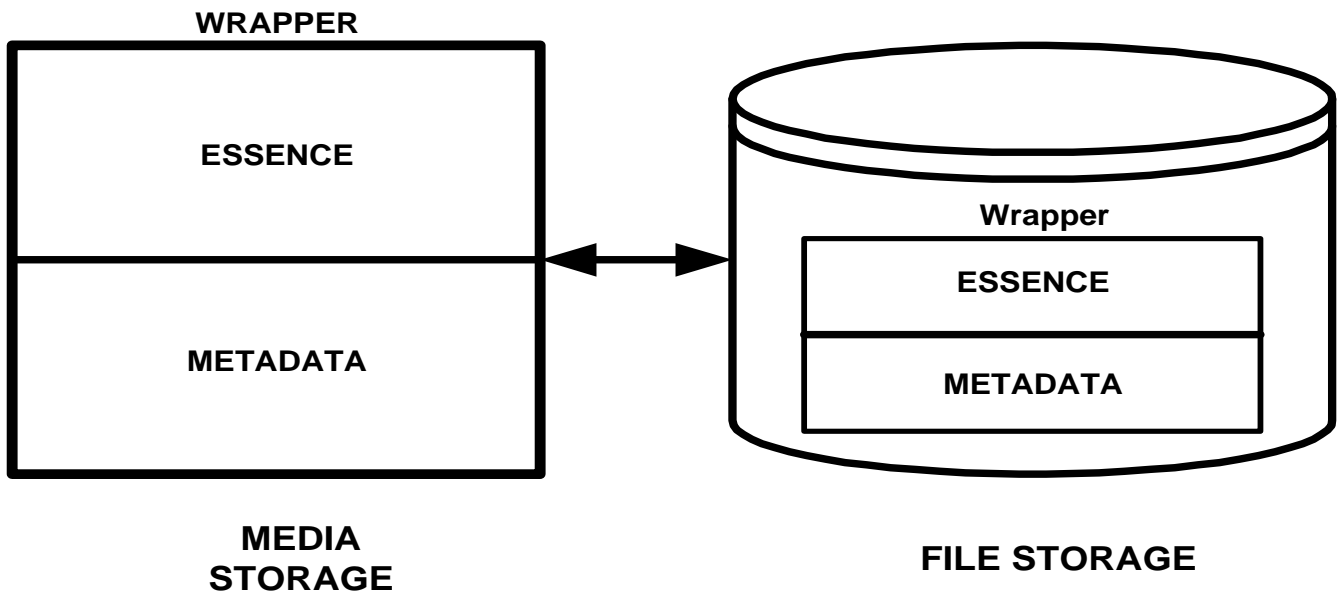
There is no inherent limitation in the range of Metadata in Content, although as a practical matter the Model identifies eight categories, only three of which can be generalized to non-video and audio Essences.  The three that can be generalized include Essential Metadata, which is information essential to decode the Essence as in the case of a compressed bit map image.  The second category is Access Metadata that consists of information regarding access rights information, such as confidentiality. The third category is Descriptive Metadata that encompasses all the information used in the description, search, retrieval, and management of Content.  This can include labels, author, location, version information, and transaction records.

*4.2.3.2   Wrapper and Metadata Requirements.*

The Task Force Report identifies several specific Wrapper and Metadata requirements that include:

- **Platform neutrality**.  This means that Wrappers must be capable of being read by any computer without regard for the computer used originally to create the Wrapper.

- **Compatibility**. That Wrapper formats must be compatible with existing formats for Essence and established working practices.   A Wrapper format is considered compatible when Content (Essence and Metadata) from a source can be directly placed in a Wrapper or directly imported from another Wrapper.

- **Extensibility**.  Wrapper formats should be expected to have a reasonable longevity of decades or more.  This is especially important because over time it is likely that new Metadata types and Essence formats will be implemented as new standards evolve. Therefore every Wrapper format should be extensible in terms of being able to accommodate new Essence and Metadata types.

- **Interoperability.**  Wrapper formats must be capable of supporting the transfer and decoding of any and all Essence and Metadata type.  In addition, Wrapper formats must be capable of interpreting Complex Content, that is multiple Essence instances with many interrelationships.

- **History**.  This identifies two types of historical information that can be included in the Metadata.  Derivation history information provides historical information about editing history and signal transformation while Transaction logging tracks the steps used to produce the current version of the Content.

- **Transaction support.** This Metadata requirement includes information about copying, moving, and modification of Content.

### 4.2.3.3   Recommendations

A section in this chapter called "Framework of a Solution" lists a number of recommendations regarding Metadata.  The recommendations that are relevant for the management of electronic patent and trademark case files during their authorized retention include:

- Development of an extensible hierarchical classification of Metadata types including one or more Metadata templates that are appropriate for particular users

- Establishment of a registry of Metadata types and definitions

- Standardization of a generic Wrapper for the Storage of Content

## 4.3   SUMMARY

This section has provided a definition of metadata and reviewed five different perspectives on the definition and use of metadata with electronic resource material that can be applied to the life cycle management of electronic patent and trademark case files.  These perspectives provided insight into the requirements for and methods used to help facilitate the storage and retrieval of case files and also to support their interoperability and exchange among and between heterogeneous computer platforms as technology changes.  The next section will discuss several broad metadata considerations for the management of electronic patent and trademark case files that are underpinned by these five perspectives.

# 5  METADATA CONSIDERATIONS

This section explores three areas of metadata considerations that support the long-term access and retention of electronic patent and trademark case files. These considerations take into account the regulations and guidelines, and the findings and recommendations from the studies presented and discussed in Section ??. The three implication areas are: usable, reliable, and "wrappers."

## 5.1  USABLE

Code of Federal Regulations (CFR1234.24(b)(1)(ii) requires agencies to "Retain records in a usable format for their required retention period . . . ."  There are few technical problems that agencies encounter when complying with this requirement so far as paper-based records are concerned largely because paper-based records are self-referential in the sense that they are eye-readable without the intervention of a technology and that the accuracy of their intellectual content and their trustworthiness as records could be inferred from the records themselves.  Thus, complete paper-based records inherently are usable because they are understandable and organized in some systematic way.

 This level of assurance about the usability of paper-based records is not readily extended to electronic records that exist as logical entities and must be interpreted by software and then rendered into an intellectual form that humans can understand and use.  The intervention of software to render electronic records for human use gives rise to an inherent software dependency that is unavoidable and inevitable.  Furthermore, the logical structure of the electronic records, especially those that combine different formats (e.g., text and image), may be physically separated so that links and pointers have to be maintained in order to render the records for human use.

The information that is essential for retaining the usability of patent and trademark case files over their authorized disposition period by, especially as technology changes, is achieved through the capture and maintenance of history metadata. Specifically, this means that the information at both the case file level and the discrete record level that preserves the integrity of the electronic record structure, content, and context and enables the records to be rendered as intelligible and understandable to humans must be captured and maintained. For patent and trademark case files, the collection and maintenance of appropriate case file and record level metadata can be used to facilitate the retrieval and rendering (display or printing) of the content of an entire case file or a single record that is intelligible and understandable.

## 5.2 INTEGRITY AND AUTHENTICITY

Protecting the integrity and, therefore, the authenticity, of patent and trademark case files is a key legal (and credibility) requirement for the PTO. Hence, metadata that can help ensure the integrity of patent and trademark case files and the individual records comprising them is a fundamental requirement. A key aspect of this metadata is its capacity to document that when the underlying bit streams of electronic records that comprise patent and trademark case files are exposed to the potential risk of loss or alteration during media renewal or transfer to a new technology platform, there is no resulting alteration or loss in their structure, content, and context. This documentation comprises an "audit trail" that can be used to identify and reconstruct the results of all activities that have a potential risk exposure.

## 5.3 WRAPPERS

The "case file wrapper" now used with paper-based patent and trademark processing and maintenance serves several functions. First, it physically distinguishes a case file from other filed material. Its size, construction, and label conventions clearly denote that it is a patent or trademark case file. Second, in the case of a pending patent or trademark the wrapper identifies its contents as being highly confidential and therefore it must be protected from unauthorized access or viewing at all times. Third, the "case file wrapper" is used to physically arrange the

content according to a prescribed scheme that facilitates quick identification of the location of a specific type of record.

An "electronic wrapper" for electronic patent and trademark case files shares some of these same functions. For example, an electronic case file wrapper would distinguish it from other electronic information objects and also denote, where appropriate, that the confidential nature of its contents must be protected. However, the logical entities that comprise the records in an electronic patent or trademark case file are not inherently "self-referential" and therefore require explanatory and linkage information along with the information content.

The "Wrapper Model" proposed in the SMPTE Task Force Report discussed earlier provides useful guidance that can be extrapolated to patent and trademark case files. The proposed "Wrapper Model" contains "content" and also defines and describes the structure of the content. "Content," therefore, is divided into "Essence," that is, the actual signals that comprise the information content, and "Metadata," that is, information that describes the "Essence" and facilitates its correct interpretation. Only two Metadata categories of this 'Wrapper Model" are relevant for this study. The first is called "Essential Metadata," which means information that is necessary to decode the information content of "Essence," as in the case of compressed bit map images or encrypted records. The second is called "Descriptive Metadata" and it includes all information used in the description, search, retrieval, and management of the Content of a "Wrapper." The metadata requirements to support long-term support of patent and trademark case files discussed in the next section generally take into account the approach in this model.

### 5.3.1 Logical vs. Physical Encapsulation

The "Wrapper Model" also could provide the function of encapsulating or integrating all of the content and metadata information that comprises a patent or trademark case file into a single digital object. This encapsulation or integration could be either as a logical or physical entity.

Logical encapsulation would mean that all of the information comprising the metadata and content would be linked together but potentially stored on different file servers and possibly on different volumes of media. For example, XML records might be stored on an XML file server, bit map records might be stored on an image file server, and vector drawing records might be

stored on a vector file server. In some instances, the metadata at the "Wrapper-level" might contain pointers to metadata stored in one or more individual file servers or volumes of media.

Logical encapsulation minimizes centralized storage requirements for metadata associated with each case file. The disadvantage of logical encapsulation is that it involves very complex configuration management to minimize the loss or corruption of pointers that decouple elements of a record or a case file. This could happen when a file server is updated and the links or pointers from other file servers to a lookup table no longer work or a particular file server is taken out of service and its contents transferred to another file server without also updating the links and pointers.

Physical encapsulation, which works best with removable storage media, involves placing all of the metadata and content information within a "Wrapper" on a single storage medium. Physical encapsulation may provide a self-referential element for patent and trademark case files because all of the metadata and content, including all file formats, of a case file would be physically stored as a single entity.

Physical encapsulation, of course, minimizes the risk of pointers and linkages being corrupted that would result in the inability to fully reconstitute the entire case file at some subsequent point in time. Physical encapsulation is not without its own problems. For example, suppose that the PTO introduces a new technology platform that supports new file formats that have backward compatibility to older file formats. The transfer of records in the old file format to the new one could involve considerable processing time just to identify all of the physically encapsulated patent and trademark case files that contain records in the old format. The required time simply to identify the case files containing records in the old format could be minimized if the metadata for each case file record also identifies the file formats used to store and process the records it contains.

The decision about logical or physical encapsulation of electronic patent and trademark case files must be made in the context of the functionality of the planned AIS, and the degree to which the management of  case files might degrade the AIS performance or otherwise diminish cost-effectiveness.  In short, in the current environment the question of logical or physical

encapsulation of patent and trademark case files should be decided based on a study of the

technical  and economic impact of each method.

# 6  METADATA REQUIREMENTS

## 6.1  OVERVIEW

This section proposes metadata requirements to support the long-term retention of electronic patent and trademark case files. It incorporates requirements from the Code of Federal Regulations (CFR) and synthesizes those identified in DOD 5015.2, the University of Pittsburgh "Requirements for Business Acceptable Communications," and the SMPTE Task Force on Harmonized Standards.

These requirements are organized into six categories: Case File Profile, Case File Record Profile, Use History Profile, Reformat Profile, Copy Profile, and Transfer Profile. The Reformat, Copy, and Transfer Profiles are based upon the definitions established in the report produced for Task Order 98-03, Deliverable 98-03-10 *Long Term Access and Migration Strategy,* December 24, 1998.

These Metadata Profiles do not identify how each should be captured or when it should be done. Nonetheless, it assumes that some metadata elements, such as those for the Case File Profile and Case File Record Profile can be automatically acquired from the updated PALM and TRAM databases. Many of the other metadata elements, such as Access Dates, are assumed to be supplied automatically by the AIS or, as a last resort, via manual data entry. Another important feature of these profiles is the linkage of individual data elements within and between profiles. For example, the Use History Profile contains a metadata element called "Access Dates" that is a sequential date list of every access to a case file. These "Access Dates" could be linked to the dates when a specific record was accessed or to the dates when the entire case file or individual records were reformatted, copied, or transferred.

Refinements to the elements and properties of these metadata profiles will be more explicitly defined and refined, particularly in terms of their relationship to PALM and TRAM upgrades and

to other electronic patent and trademark AISs and databases based on comments by and reviews with the PTO.

The following decribes the six metadata requirements profiles.


## 6.2   CASE FILE PROFILE

The data elements comprising the Case File Profile capture core information about the case file as an "encapsulated" logical entity that supports long-term retention and management.  All of these data elements are considered non-revisable from the moment that they are registered as part of the case file. The only instance of a change that could occur in the casefile profile is when there are updates to the case file after closure, such as assignment of rights or reexamination, or if the location status changes, or access dates are transferred from the individual record profile.

**CASE FILE METADATA PROFILE**

**Case File Identifier**

**Case File Type**

      **Patent**

      **Trademark**

**Case File Subject**

**Filing Date**

**Closure Date (defines the beginning of the authorized disposition period)**

**Process Status Code (pending, issued, registered, abandoned)**

**Representation**

      **Binary**

      **ASCII**

      **EBCDIC**

      **UNICODE**

**Encryption of Case File (optional, if required)**

      **Name of Algorithm Used**

      **Name of Software Used**

**Formats Used**

      **Text**

      **Image**

      **Vector**

      **Compound (e.g. text and vector combined)**

      **Database**

      **Audio**

      **Moving Image**

**File Formats Used**

      **TIFF**

      **XML**

      **SVG**

      **JPEG**

      **CNG**

      **MPEG**

      **Other\***

**Case File Size**

      **Logical Record Length**

      **Logical Record Count**

      **Physical Record Count**

      **Bye Count**

      **Total Number of Records**

**Case File Authentication (If Used)**

      **CRC**

      **Hash Digest**

**Updates After Closure\***

      **Assignment**

      **Contested**

**Access Dates List\***

**Location of Case File\***

      **PTO Operations**

      **PTO Repository**

**Federal Records Center**

**Related Paper or Microfilm Records**

**Disposition \***

      **Instruction Code**

      **Action Date**

**\*Indicates update to the profile is permitted**

## 6.3   CASE FILE RECORD PROFILE

The data elements comprising the case file record profile capture core information about each discrete record in the case file that can be used individually or collectively to provide intellectual control of the records.  For example, the Access Dates data element contains all instances of access action to the record.  A specific date can be linked to the Use History Metadata Profile, for example, that can would disclose if the access was linked to an update, reformat, copy, or transfer activity.  Each of these data elements is considered non-revisable from the moment the record becomes part of a case file. The only instance of a change that could occur in the case file record profile is when access dates are added.  An asterisk ("\*") identifies these instances.

**CASE FILE RECORD METADATA PROFILE**

**Case File Unique Identifier**

**Unique Record Identifier (e.g., serial number)**

**Record Descriptor (form number, alpha code)**

**Record Receipt/Creation Date**

**Name of Record Recipient/Addressee**

**Subject**

**Originating Organization**

**Author**

**Security Level (Pending, Classified)**

**Representation**

> **Binary**

> **ASCII**

> **EBCDIC**

> **UNICODE**

**Encryption of Record (optional, if required)**

> **Name of Algorithm Used**

> **Name of Software Used**

**Formats Used**

> **Text**

> **Image**

> **Vector**

> **Compound (e.g., text and vector)**

> **Database**

> **Audio**

> **Moving Image**

**File Formats Used**

> **TIFF**

> **XML**

> **SVG**

> **JPEG**

> **CNG**

> **MPEG**

> **Other\***

**Record Size in Bytes**

**Record Authentication (If Used)**

> **CRC**

> **Hash Digest**

**Access Dates\***

**\*Denotes that updates are permitted**

## 6.4   USE HISTORY METADATA PROFILE

The Use History Metadata Profile is intended to track information about activities and events related to the electronic case file after closure (post issuance, registration or abandonment).  It is assumed that prior to closure, the upgraded PALM and TRAM or associated work flow systems, will provide control, tracking and audit trail capabilities.

The purpose of the Use History Metadata Profile is two fold: 1) to establish basic information about each case file from the time of closure, and 2) to build an audit trail that documents when the case file was accessed and what processing actions were taken.  It is assumed that over the authorized retention life of patent and trademark case files that there will be multiple accesses, reformats, copies, and transfers.  The iteration number for reformat, copy, and transfer serve as direct links to more detailed records.  The Date profile element for Reformat, Copy, and Transfer could be linked to the Access Dates data element to the Case File Record Metadata Profile to distinguish between these actions and mere access for information purposes.  There is an optional Purge/Delete that can be invoked when a case file is not selected for transfer to NARA (as it likely to be the case with some trademark case files).  Thus, basic information about every patent and trademark case file the PTO opens will be permanently retained  even when the case file itself may be purged or deleted .

### USE HISTORY METADATA PROFILE


**Case File Unique Identifier**

**Event/Activity Date (Repeatable)**

**Access Dates (Repeatable)**

**Reformat  (Repeatable)**

       **Date**

        **Reformat Iteration Number**

        **Logical Record Size**

        **Logical Record Count**

        **Byte Count**

        **CRC (If Used)**

        **Hash Digest (If Used)**

    **Copy (Repeatable)**

        **Date**

        **Copy Iteration Number**

        **Logical Record Count**

        **Byte Count**

        **CRC (If Used)**

        **Hash Digest (If Used)**

    **Transfer (Repeatable)**

        **Date**

        **Transfer Iteration Number**

        **Logical Record Count**

        **Byte Count**

        **CRC (If Used)**

        **Hash Digest (If Used)**

    **Purge/Delete (Optional)**

        **Date**

        **Authorization**

## 6.5   REFORMAT METADATA PROFILE

This metadata profile is divided into input metadata and output metadata elements.  The purpose of these metadata elements is to capture detailed information about status of a case file before and after reformatting in order to establish the foundation for its trustworthiness over time.  At the time of the first reformatting, many of the input data elements most likely would be extracted

from the Case File or Case File Records Metadata Profiles.  Subsequent reformattings would link
backward to the most recent processing, which could be reformatting, copying, or transferring,
and extract the relevant metadata elements.

### REFORMAT METADATA PROFILE (INPUT)

**Date of Reformatting**

**Reformat Iteration Number**

**Case File Identifier**

**Case File Record Identifier (If Appropriate)**

**File Formats Used**

> **TIFF**
>
> **XML**
>
> **SVG**
>
> **JPEG**
>
> **CNG**
>
> **MPEG**
>
> **Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

> **CRC**
>
> **Hash Digest**

**Storage Media**

> **Vendor**
>
> **Type (e.g., RAID, 3480 or DLT Tape)**
>
> **Product Name**
>
> **Volume ID**

**Software Used in Reformatting**

> **Name Of Product**
>
> **Version Number**

## REFORMAT METADATA PROFILE (OUTPUT)

**Date of Reformatting**

**Reformat Iteration Number**

**Case File Identifier**

**Case File Record Identifier (If Appropriate)**

**File Formats Used**

> **TIFF**
>
> **XML**
>
> **SVG**
>
> **JPEG**
>
> **CNG**
>
> **MPEG**
>
> **Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

> **CRC**
>
> **Hash Digest**

**Storage Media**

> **Vendor**
>
> **Type (e.g., RAID, 3480 or DLT Tape)**
>
> **Product Name**
>
> **Volume ID**

**Comparison**

> **Byte Count**
>
> **CRC**
>
> **Hash Digest**
>
> **Visual Inspection**
>
> **Discrepancies (If Any)**
>
> **Corrections (If Any and Explanations)**
>
> **Supervisor Review**

**Physical Storage Location**

      **Primary**

      **Backup**

## 6.6   COPY METADATA PROFILE

The purpose of the metadata elements in the copy format template is to capture detailed information about status of a case file before and after copying in order to establish the foundation for its trustworthiness over time.  Consequently, the Copy Metadata Requirements are divided into two categories -Input and Output.  Copying patent and trademark case files can occur at the time after closure, after initial reformatting, or in conjunction with transfer (discussed later).  At the time of the first copying many of the input data elements most likely would be extracted from the Creation-Use or Reformat Template.  Subsequent copying would link backward to the most recent processing, which could be reformatting, copying, or transferring, and extract the relevant metadata elements.

<div align="center">

**COPY METADATA PROFILE (INPUT)**

</div>

**Date of Copying**
**Copy Iteration Number**
**Case File Identifier**
**Case File Record Identifier (If Appropriate)**
**File Formats Used**
      **TIFF**
      **XML**
      **SVG**
      **JPEG**
      **CNG**
      **MPEG**

**Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

      **CRC**

      **Hash Digest**

**Storage Media**

      **Vendor**

      **Type (e.g., RAID, 3480 or DLT Tape)**

      **Product Name**

      **Volume ID**

**Software Used in Reformatting**

      **Name Of Product**

      **Version Number**


## COPY METADATA PROFILE (OUTPUT)


**Date of Copying**

**Copy Iteration Number**

**Case File Identifier**

**Case File Record Identifier (If Appropriate)**

**File Formats Used**

      **TIFF**

      **XML**

      **SVG**

      **JPEG**

      **CNG**

      **MPEG**

      **Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

      **CRC**

**Hash Digest**

**Storage Media**

    **Vendor**

    **Type (e.g., RAID, 3480 or DLT Tape)**

    **Product Name**

    **Volume ID**

**Comparison**

    **Byte Count**

    **CRC**

    **Hash Digest**

    **Visual Inspection**

    **Discrepancies (If Any)**

    **Corrections (If Any and Explanations)**

    **Supervisor Review**

**Physical Storage Location**

    **Primary**

    **Backup**

## 6.7   TRANSFER METADATA PROFILE

Like the Reformat and Copy Metadata Requirements, the purpose of the Transfer Metadata Profile is to capture information that documents fully the actions taken in this activity that will help support the trustworthiness of electronic patent and trademark case files despite changes in technology.  The metadata elements in this profile capture detailed information about status of a case file before and after transfer.  It is likely that at the time of the first transfer many of the input data elements most likely would be extracted from either the Reformat or Copy Templates.  Subsequent transfers would link backward to the most recent processing, which could be reformatting, copying, or transferring, and extract the relevant metadata elements.

## TRANSFER METADATA PROFILE (INPUT)

**Date of Transfer**

**Transfer Iteration Number**

**Case File Identifier**

**Case File Record Identifier (If Appropriate)**

**File Formats Used**

> **TIFF**
>
> **XML**
>
> **SVG**
>
> **JPEG**
>
> **CNG**
>
> **MPEG**
>
> **Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

> **CRC**
>
> **Hash Digest**

**Storage Media**

> **Vendor**
>
> **Type (e.g., RAID, 3480 or DLT Tape)**
>
> **Product Name**
>
> **Volume ID**

**Software Used in Transferring**

> **Name Of Product**
>
> **Version Number**

## TRANSFER METADATA PROFILE (OUTPUT)

**Date of Transfer**

**Transfer Iteration Number**

**Case File Identifier**

**Case File Record Identifier (If Appropriate)**

**File Formats Used**

> **TIFF**
>
> **XML**
>
> **SVG**
>
> **JPEG**
>
> **CNG**
>
> **MPEG**
>
> **Other**

**Case File or Record Byte Count**

**Record Authentication (If Used)**

> **CRC**
>
> **Hash Digest**

**Storage Media**

> **Vendor**
>
> **Type (e.g., RAID, 3480 or DLT Tape)**
>
> **Product Name**
>
> **Volume ID**

**Comparison**

> **Byte Count**
>
> **CRC**
>
> **Hash Digest**
>
> **Visual Inspection**
>
> **Discrepancies (If Any)**
>
> **Corrections (If Any and Explanations)**

**Supervisor Review**

**Physical Storage Location**

**Primary**

**Backup**

# 7 RECOMMENDATIONS

This section proposes three recommendations for consideration by the PTO to implement that will help support the long-term management and use of electronic patent and trademark case files. Specifically, these three recommendations can help provide the assurance that patent and trademark case files will be both usable and trustworthy no matter what technology changes may occur. These recommendations are derived from an analysis of metadata requirements identified in relevant sections of the CFR, DOD 5015.2, and the findings of three major metadata projects and studies. Although the three recommendations are framed by this analysis and these findings, they also take into account for broader metadata implications that include:

- **Technology platform neutrality**, which means that the metadata and case file content must be capable of being read by any computer without regard for the computer or application software used originally to create them;

- **Compatibility**, which means that the electronic case file "wrapper" formats must be compatible with existing file formats and established working practices;

- **Extensibility**, which means that the electronic case file "wrapper" format along with the metadata and record content should have a longevity of at least one decade or more and must be extensible in the sense that the "wrapper" format must be capable of accommodating new file formats and metadata types.

The three recommendations are:

## 1. Review, Refine, Integrate, and Prototype

It is recommended that the proposed metadata profiles be reviewed, refined, integrated into the overall metadata and electronic file wrapper plans and a prototype evaluation conducted.. The prototype would include conducting tests of the proposed Case File and Case File Record profiles within the context of overall patent and trademark metadata and electronic file wrapper

management and determining the effectiveness of the metadata profiles for Use History, Reformat, Copy, and Transfer.

The goal of the review, refinement process and the integration and prototype tests is to determine the level of compatibility with the current definitions and plans for an "Electronic File Wrapper" and to determine the degree of compatibility with the PALM and TRAM upgrades

Another goal of the prototype process is to test the ability of the metadata requirements to the usability and integrity preservation of electronic patent and trademark case files and all records associated with them.

Over the forty year authorized disposition period for patent case files, for example, the metadata generated under each of the six Metadata Profile categories could be quite substantial and therefore would create a significant overhead cost to the operational program.  These and other related issues should provide part of the context within which this prototype evaluation is conducted.

## 2.  Develop A Common Metadata and File Wrapper Foundation

The discussion of an electronic patent and trademark case file "wrapper" in this study is formed largely by what is happening in areas that are somewhat removed from PTO day to day operations, such as the three research studies 36 CFR regulations.  Furthermore, there appear to be several information technology programs (e.g., PALM and TRAM upgrades and TEAM) underway that are in the process of defining and developing the foundation for metadata requirements and electronic file wrapper management .  It is important, therefore, that the metadata profiles and properties and the concepts of electronic case file management as it relates to these profiles be considered in the overall definition and development of metadata and electronic file wrapper programs in both the patent and trademark automation programs.  Based on the similarities of requirements for long term access and retention of patent and trademark electronic case files, it is suggested that consideration be given to developing a common metadata and electronic file wrapper foundation that could serve the needs of both the patent application and trademark electronic work places.

### 3.   Conduct a Cost Benefit Analysis of Logical and  Physical "Encapsulation"

This study has identified several high level potential risks and benefits of employing either a physical or logical encapsulation ("case file wrappers") of electronic records contained in patent and trademark case files.  This identification of risk and benefits, however, is at an abstract level that does not take into account the specific information technology and supporting infrastructure of the PTO.  For example, it may be that the existing information technology now in place along with planned upgrades is such that only logical case file "encapsulation" makes sense either from a financial perspective or the PTO organizational culture.

The decision on the implementation of a logical or physical case file wrapper must be evaluated in a more in-depth study that is based on the costs and benefits as they relate to: a)  existing and planned technology for PTO server infrastructure, storage management strategy, b) planning and implementation of the PALM and TRAM upgrades, c) the definition and implementation direction and status of the "electronic file wrapper", and d) the definitions and recommendations in this study for metadata profiles and case file wrapper methodology for supporting long term access and retention of electronic patent and trademark case file records.