A Formal Method for Analyzing the Authenticity Properties of Procedures for Preserving Digital Records

William E. Underwood
Georgia Tech Research Institute
Atlanta, Georgia, 30332-0832 USA
william.underwood@gtri.gatech.edu

Abstract. A formal method is described for analyzing records management and archival procedures and systems to determine whether they maintain and preserve authentic records over time. The analysis procedure is based on a formalization of archival and diplomatic concepts and principles as definitions and axioms. Concepts such as digital record, record series, and archival integrity are defined and axioms characterizing authentic documents and authentic records are formulated. A procedure is described for storing and retrieving the digital records of a record creator that incorporates elements to ensure the integrity and authenticity of the records. The theories of record integrity and authenticity are used with theories of communications security and belief to prove that the procedure achieves its goal of preserving the integrity and authenticity of the digital records. This demonstrates the formal method of analysis.

Keywords: record integrity, record authenticity, digital preservation

1. Introduction

The rapid obsolescence of computing technologies creates difficulties for those concerned with the long-term preservation of records in digital form. The potential need to migrate these records across hardware and software technologies raises questions related to the records' authenticity. How can one ensure that sets of digital records have not been intentionally or inadvertently modified? How can one ensure that long-term preservation methods do not compromise the authenticity of digital records?

The research reported in this paper is exploring the use of a formal method for analyzing records management and archival procedures and systems to determine whether they maintain and preserve authentic records. The analysis procedure is based on a formalization of archival and diplomatic concepts and principles.[1]

[1] This paper was completed as part of the InterPARES Research Project (http://www.interpares.org). This project, which is investigating the preservation of permanent authentic records in electronic record-keeping systems, brings together an interdisciplinary research team drawn from National Archives and universities in North America, Europe, and Asia. The project has received substantial support from the Canadian Social Science and Humanities Research Council Major Collaborative Research Initiative (MCRI), the National Historical Publications and Research Commission in the United States, and the Italian National Research Council. The author of this paper is a member of the American InterPARES Research Team and the InterPARES Preservation Task Force. This research was also supported in part by the Electronic Records Archives Program of the National Archives and Records Administration.

In the next section, diplomatic concepts such as document, document content, documentary form, competence, archived document (record), record integrity, and record authenticity are represented in a formal, logical language. These concepts are extended to include digital documents and digital records. Archival concepts such as file and record series are also formalized.

In the third section, a procedure for archiving digital records and verifying their authenticity is described. In section 4, a method for proving the correctness of procedures is described and is used to prove the correctness of procedures for maintaining and verifying the integrity and authenticity of digital records. In the final section, additional research issues are identified.

2. Axiomatic Theories for the Analysis of Record Authenticity

Diplomatics provides a theory of documents that can be used to determine whether a document is authentic.[2] Archival Science provides a theory of aggregations of records.[3] In this section, some of the fundamental concepts of these sciences are represented in terms of set theory.

The axiomatic method involves writing down in a formally specified language what is known about a domain of discourse; defining the rules by which other statements can be derived; then identifying those statements in the language from which the other statements can be derived. In the full technical report, the formal language used to express these theories is defined.[4] Verification of the authenticity of electronic records that are stored in distributed computing systems requires the capability to reason about communications protocols and authentication mechanisms. In the full report, a logical theory of communications is described that has been used to analyze the authentication and confidentiality properties of communications protocols. Verification of the authenticity of electronic records stored in computing systems cannot be on the basis of complete information. Rather, it involves assumptions about the trustworthiness of record creating agents, record keeping systems and archival systems. It requires assumptions about shared secrets and authentic private keys used for digital signatures. In the full report, a logic of belief is described that allows one to reason and reach conclusions based on incomplete knowledge.

2.1 Documents

A document is information consigned to a medium.[5] The *medium* of a record is the physical carrier of the content and form of the record.[6] The concept of medium as

[2] L. Duranti. *Diplomatics: New Uses for an Old Science*. Lanham, Maryland: Scarecrow Press, 1998.

[3] L. Duranti. "Archival Science," *Encyclopedia of Library and Information Science*, Allen Kent ed., vol. 59 (New York, Basel, Hong Kong: Marcel Dekker, Inc., 1996), 1-19.

[4] W. E. Underwood. A step towards a logical theory of record integrity and authenticity, US-InterPARES Technical Report, Information and Telecommunications Laboratory, Georgia Tech Research Institute, Atlanta Georgia, March 2002. http://is.gseis.ucla.edu/us-interpares/

[5] UBC Project, Glossary, March 1997.

used in Diplomatics and Archival Science can be interpreted as being a communications channel. A communication channel is a means of conveying information from one principal to another. Communication channels such as paper, display screens, tapes, and disks are called communication channels with memory.

A document is a message that is quoted by a channel that speaks for a principal.

Axiom A1: $Document(X) \supset Channel(Q) \wedge P\ states\ Q\ speaks\ for\ P \wedge Q\ quotes\ P\ states\ X$

A principal *P*'s signature on paper document *X* amounts to a statement that the document speaks for *P*. The channel (medium) quotes the principal *P* as stating *X*.

The *content* of a record is the facts or information in the record.[7] The semantic content of a document is the meaning of the terms and statements in the document. The syntactic content of a document is the terms (names of principals, keys, propositions, dates, graphics, and images) of the document, but not the functions of terms such as font, font size and layout.

Definition 1: (*syntactic*) *content of document*(X) $\equiv <x_1, \ldots, x_n>$ where x_i are terms, but not functions of terms.

When a person makes a document, he intends to make a document for some purpose, e.g., a contract, a job application, a job promotion, a progress report. A document that is made for a certain purpose has certain attributes that are essential to the intended purpose. These attributes are reflected not only in the content, but in the form of the document.

Documentary form is "The rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated."[8] Documentary form consists of extrinsic (or physical) elements and intrinsic (or intellectual) elements of form. An *extrinsic element* is "an element of a record that constitutes its external appearance. The types of extrinsic elements include presentation features, electronic signatures, electronic seals, digital time-stamps issued by a trusted third party, and special signs."[9] *Intrinsic elements* are "the elements of a record that constitute its internal composition. The types of intrinsic elements include name of author, name of originator, chronological date, name of place of origin of record, name of addressee(s), name of receiver(s), indication of action (matter), name of writer, corroboration, attestation, and qualification of signature."[10]

The form of a document is a sequence of functions defining the physical and intellectual elements of the document.

[6] ibid.

[7] L. Duranti and T. Eastwood. Protecting electronic evidence: A progress report on a research study and its methodology. *Archivi & Computer* (3) 1995 pp. 213-250, (p.222).

[8] InterPARES Project, The InterPARES Glossary

[9] ibid

[10] ibid

Definition 2: *documentary form*$(X) \equiv \langle g_1 f_1(X), \ldots, g_n f_m(X) \rangle$ where f_i are functions defining the physical elements and g_j are functions defining the intellectual elements of form.

More generally, documentary form can be defined using regular expressions, regular grammars or style sheets defining mandatory, optional and repetitive elements.

A *digital object* is an instance of an abstract data type whose principal components are data and key-metadata. The key-metadata includes a *handle*, i.e., a persistent identifier globally unique to the digital object.[11] The data of each digital object is typed. Data types may include MS Word 7, Tiff, XML, set-of-data-types, and set of digital objects. Methods are associated with the data types for such purposes as creating or viewing the data.

Data types and key-metadata can be used to create subtypes of digital objects. For instance, one could create a subtype of digital object called XML digital documents that has XML as the data type of the data. One could also create a subtype technical report in which an XML DTD represented the documentary form for a technical report.

Let $\{0, 1\}^*$ denote the set of all bit strings, i.e., sequences of 0's and 1's. A digital document is a document for which there is a function (method) C_i that encodes the document X as a $Y \in \{0,1\}^*$ and an inverse function C_i^{-1} that decodes Y to display the document X. C_i and C_i^{-1} are programs that operate on computers with storage devices for storing the bit strings produced by C_i and display devices for displaying the bit strings decoded by C_i^{-1} as documents.

Definition 3: *Digital document*$(X) \equiv$ *Document*$(X) \wedge$ *Digital object*$(X) \wedge C_i(X) = Y \wedge Y \in \{0,1\}^* \wedge C_i^{-1}(Y) = X$

A digital document X is *reproducible* (or *viewable*) if for each program $C_i(X)$ that encodes a document as a bit string Y there exists a program $C_i^{-1}(Y)$ for decoding and displaying the document X and a computer, operating system and a display device on which the program C_i^{-1} executes.

2.2 A Principal's Competence and Reliability

Archival Science and Diplomatics are based on knowledge of juridical systems and organizations. The *juridical-administrative context* of a record (or record series or archival fonds) is "the legal and organizational system in which the creating body belongs."[12] It is beyond the scope of this paper to present theories of juridical systems and organizations. However, formulae for some organizational concepts that are needed for reasoning about the reliability of statements made by principals will be presented.

[11] R. Kahn and R. Wilensky. A framework for distributed digital object services, Technical Note 95-01, Corporation for National Research Initiatives, May, 1995.

[12] The InterPARES Glossary.

Organizational goals are achieved by executing activities. Activities are sequences of actions leading to a goal.[13] *Goal-activity*(G, A) is the set of activities for achieving goal G. Organizational units have organizational-positions, e.g., director, project director, research scientist, that principals fill. *Organization-position*(O, T) is the set of organizational positions. [14] *Position-role*(T, R) is the set of roles associated with position T. *Fills*(P, T) means that principal P fills position T. *Has goal*(R, G) is the set of goals entrusted to a role. *Role-activity*(R, A) is the set of activities that role R is authorized or empowered to perform.

Competence is "the sphere of functional responsibility entrusted to an office." [15]

Axiom A3: (*Organizational-position*(O, T) $\wedge$ *Fills*(P, T) $\wedge$ *Position-role*(T, R) $\wedge$ *Has-goal*(R, G) $\wedge$ *Role-activity*(R, A)) $\supset$ *P has competence R*

An honest principal only says what he believes.

Definition 4: *Honest*(P) $\equiv$ *P states X* $\supset$ *P believes X*

By *P is a trusted authority on* φ is meant P has competence R, φ is a formula that P might state while acting in role R, and P is honest. This is a way of saying that a principal is reliable with regard to statements φ.[16]

Definition 5: *P trusted authority on* $\varphi \equiv$ *P has competence R* $\wedge$ *Honest*(P) $\wedge$ *P acting in role R states* φ

If a principal is a trusted authority on a formula, and states that formula, the formula is true.

Axiom A4: (*P trusted authority on* $\varphi \wedge$ *P states* φ) $\supset \varphi$

2.3 Records and Record Series

A *record* is a document made or received and set aside as a record in the course of a practical business activity.[17]

Axiom A5: (P, $Q \in O \wedge$ (*P acting in role R states X* $\vee$ *P acting in role R receives X*) $\wedge$ *Role-activity*(R, A) $\wedge$ *Document*(X) $\wedge$ *Q trusted authority on store X* $\wedge$ *Q stores X*) $\supset$ *Record*(X) $\wedge$ *Record of Activity*(X, A)

[13] *Activity*: a sequence of actions directed toward the achievement of one purpose. InterPARES Glossary

[14] Compare *Competent Person*: The office which is given a competence and has, therefore, the authority and capacity to act within it. The InterPARES Glossary.

[15] The InterPARES Glossary.

[16] Author's reliability: the competence of the author to issue the specific document and/or the degree to which an author can be trusted. L. Duranti and T. Eastwood, The preservation of the integrity of electronic records, Template 3: What is a reliable record in the traditional environment?

[17] ibid

In axiom A5, P could be identical with Q.

Digital records are records that are digital documents. [18]

Definition 6: $\mathit{Digital\ record}\,(X) \equiv \mathit{Record}(X) \wedge \mathit{Digital\ document}(X)$.

A *stored digital object* is a digital object stored in a repository. A *registered digital object* is a stored digital object whose handle has been registered. The initial repository used to deposit a registered digital object is designated the *repository of record*.[19]

Records are collected into logical units called files. Files enable one to refer to a set of records by name. There is some formula $\varphi(X)$ defining membership of the records in the file in terms of their subject, activity, or transaction. The records within a file are ordered according to attributes f of the records.[20]

Definition 7: $\mathit{File}\,(F) \equiv \{X\colon \mathit{Record}(X) \wedge \varphi(X) \wedge f{:}X \to Y$
$\wedge\ \rho$ is an order relation on $Y \times Y\}$

Examples: A chronological file is a file of records ordered by chronological date. An alphabetic correspondence file is a file of records that are correspondence that are ordered alphabetically by the names of correspondents.

A *record series* is files or documents arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation or filing process, the same function, or the same activity; have a particular form; or because of some other relationship arising out of their creation, receipt, or use.[21]

A record series is a set of files that contain records of a business activity.

Definition 8: $\mathit{Record\ series}(S) \equiv \{F\colon \mathit{File}(F) \ \wedge X \in F \wedge \mathit{Organization\text{-}goal}(O,G)$
$\wedge\ \mathit{Goal\text{-}activity}(O, A) \wedge \mathit{Record\ of\ activity}(X, A)\}$

Example: The "Files of the Office of Policy Development" is the title of a record series. The record series consists of a Chronological File, an Alphabetic Correspondence File, and Subject Files.

The archival fonds of an organization is the union of all record series containing records of business activities related to the organization's goals. [22]

Definition 9: $\mathit{Archival\ fonds}\ (\mathsf{A}, O) = \ \cup S_{\mathrm{i}}$ for all $G \in \mathit{Organization\text{-}goal}(O,G)$
and $A \in \mathit{Goal\text{-}activity}(G, A) \wedge \mathit{Record\ series}(S_{\mathrm{i}})$

[18] "An electronic record is any record that is made or received and initially set aside in electronic form." Duranti and Eastwood, Protecting electronic evidence.

[19] Kahn and Wilensky. A framework for distributed digital object services.

[20] *File* 1. An organized unit (folder, volume, etc.) of documents grouped together either for current use or in the process of archival arrangement. Also called a *file unit*. Bellardo and Bellardo. *A Glossary for Archivists*.

[21] Bellardo and Bellardo. *A Glossary for Archivists*.

[22] *archival fonds*: the whole of the records of a creator. The InterPARES Glossary.

2.4 Record, File and Record Series Integrity

Definition 10: *Content integrity* (or *information integrity*) is the property whereby the content of a message has not been altered since the time it was created, transmitted, or stored by an authorized source.

Definition 11: *Form integrity* is the property whereby the documentary form of a document has not been altered since the time it was created, transmitted, or stored by an authorized source.

Definition 12: *Document* (*Record*) *integrity* is the property of a document (record) whereby the content and documentary form of the document (record) have not been altered in an unauthorized manner since the time the document was made or transmitted (stored) by an authorized source.[23]

Definition 13: *File integrity* is the property of a file whereby each of the records in the file has record integrity and there have been no unauthorized insertions, deletions, substitutions or reordering of records in the file since the times of creation of the records.[24]

Definition 14: *Record series integrity* is the property of a record series whereby each of the files in the record series has file integrity and there have been no unauthorized insertions, deletions, substitutions or reordering of files in the series since the files were created.

Definition 15: *Archival integrity* is the property of an archival fonds whereby each of the record series in an archival fonds have record series integrity and there have been no unauthorized insertions, deletions, substitutions or reordering of record series since the time of creation of the record series.[25]

Definition 16: *Metadata integrity* is the property whereby metadata (statements about data) has not been altered in an unauthorized manner since the time it was created, transmitted, or stored by an authorized source.

2.5 Record Authenticity

Authentication may be informally defined as the process of verifying that the identity of the source of an object is as claimed.

[23] Compare "The *integrity of a record* is its wholeness and soundness–that it is intact and uncorrupted." InterPARES Project, Draft Requirements for Authenticity. p. 4.

[24] File integrity: The concept that the accuracy, completeness, and original order of the records in a filing system must be maintained. Bellardo and Bellardo, *A Glossary for Archivists*.

[25] Archival integrity: The principle that a fonds or record group must be preserved without division, mutilation, alienation, unauthorized destruction or any addition, except by accrual or replevin, in order to ensure its full evidential and informational value. The concept of archival integrity derives from the principles of provenance and respect for original order. Belardo and Belardo, *A Glossary for Archivists*.

Definition 17: *Signature authentication* is a process (or procedure) whereby a signature is verified as being created by the specified entity.[26]

If the signature *S* affixed to a document *X* is verified to be that of the specified principal *P*, we will say *authentic signature*(*S*, *X*, *P*).

An *authentic document* is a document that is what it purports to be.[27] If *X* is a document, then what *X* purports to be is a formula $\varphi(X)$ that is implied by *X*, in other words, a conjunction of functions or properties of *X*.

If a document is written according to the practice of the time and place indicated in the text, and signed with the name(s) of the person(s) competent to create it, then it is a *diplomatically authentic document*. If a document lacks written physical or intellectual elements of form of the practice of the time and place indicated in the text, then it is diplomatically inauthentic.[28]

Axiom A7: *Document*(*X*) $\wedge$ *Document type*(*X*) = *T* $\wedge$ *signature*(*X*) = *S*
$\wedge$ *chronological date*(*X*) = *d* $\wedge$ *authentic signature*(*S*, *X*, *P*) $\wedge$ *P has competence R*
$\wedge$ *Role-activity*(*R*, *state X* $\wedge$ *Document type*(*X*) = *T*) $\supset$ *Authentic document*(*X*)

An *authentic record* is a record that is what it purports to be and is free from tampering or corruption.[29] "A record is authentic when it can be proved to be that which it is claimed to be at some point in time after its creation (whether days or centuries after its date of compilation or receipt). Proving a record's authenticity does not make it more reliable than it was when it was created. It only warrants that the record does not result from any manipulation, substitution, or falsification occurring after it has been made or received. Authenticity is provided to a record by the controls established on its transmission and preservation."[30]

Clifford Lynch observes that "Validating authenticity entails verifying claims that are associated with an object—in effect, verifying than an object is indeed with it claims to be, or what it is claimed to be (by external metadata)." "There are two basic strategies for testing a claim. The first is to believe the claim because we can verify its integrity and authenticate its source, and because we choose to trust the source."[31]

If a record and the statements of a trusted record creator about the record have not changed since they were stored, and the statements of the record creator corroborate the claims of the record, then the record is authentic.

Axiom A8: (*Record*(*X*) $\wedge$ *name of individual record creator*(*X*) = *P*

[26] "Signature authentication is the legal recognition that a signature is affixed by and belongs to the person whose name it expresses." L. Duranti, Diplomatics: New uses for an old science (Part V), *Archivaria* 32 (summer 1991), p. 9.

[27] "Document authentication is the legal recognition that a document is what it purports to be." ibid.

[28] L. Duranti. Diplomatics: New uses for an old science. *Archivaria*, 28 (summer 1989), p. 17-18.

[29] The InterPARES Glossary.

[30] Duranti and Eastwood, Protecting electronic evidence, p. 216.

[31] C. Lynch, Authenticity and integrity in the digital environment: An exploratory analysis of the central role of trust. *Authenticity in the Digital Environment*, (Washington: CLIR, 2000) pp. 32-50.

$\wedge\ \mathit{Record\ integrity}(X) \wedge \mathit{Metadata\ integrity}\varphi(X) \wedge \mathit{Consistent}(X,\ \varphi(X))$
$\supset \mathit{Authentic\ record}\ (X)$

The definition of *name of individual record creator* includes the conditions that the record creator be a trusted authority on statements about the record and that the record creator stated $\varphi(X)$. Axiom A8 does not include the condition that the document that is stored as a record must be authentic. If the record creator authored the document stored as a record, then the document would be authentic as well. However, if the records creator received the document, the authenticity of the document would need to be verified. Otherwise, one would not be able to conclude that the document was what it claimed to be. Hence, an alternative to the above axiom would be

Axiom A8': $(\mathit{Record}(X) \wedge \mathit{name\ of\ individual\ record\ creator}(X) = P$
$\wedge\ \mathit{Authentic\ document}(X)$
$\wedge\ \mathit{Record\ integrity}(X) \wedge \mathit{Metadata\ integrity}(\ \varphi(X) \wedge \mathit{Consistent}(X,\ \varphi(X))$
$\supset \mathit{Authentic\ record}(X)$

Axiom A8' implies that if the document had a digital signature, it and an authentic public key certificate of the principal would need to be stored with a record in order to verify the authenticity of the record. Alternatively, the record creator could verify the digital signature and state that the signature was authentic. The record creator's statement that the signature had been verified could be stored with the record. As a further alternative, for those records that did not have a digital signature, e.g., those that had been generated by an email system where the originator had used a key shared with the email system, the record creator could state, for inclusion in the metadata of the record, his belief as to the name of the author.

"The authenticity of a record, or rather the recognition that it has not been subject to manipulation, forgery, or substitution, entails guarantees of the maintenance of records across time and space (that is, their preservation and transmission) in terms of the provenance and integrity of records previously created."[32] "While a reliable record is one whose content you can trust, an authentic record is one whose provenance you can believe."[33] These characterizations of authenticity in terms of provenance suggest the following axiom.

Axiom A8'': $(\mathit{Record}(X) \wedge \mathit{Authentic\ document}(X) \wedge \mathit{provenance}(X) = O$
$\wedge\ \mathit{Record\ integrity}(X) \wedge \mathit{Metadata\ integrity}(\ \varphi(X) \wedge \mathit{Consistent}(X,\ \varphi(X))$
$\supset \mathit{Authentic\ record}(X)$

Axiom A8'' can be shown to be equivalent to axiom A8'.

The following definition is needed when an archivist must ensure the authenticity of a record series, that is to say, that there have been no unauthorized deletions of records in a file or of files in a record series. A record series *S* is authentic if and only if all of the records in the files of the series *S* are authentic, and the series has record series integrity.

[32] M. Guercio. Principi, metodi e strumenti per la formazione, conservazione e utilizzo dei documenti archivistici in ambiente digitale. *Archivi Per la Storia*. XII, 1-2, 1999, p. 36

[33] Duranti and Eastwood, Protecting electronic evidence, p 242.

Definition 18: *Authentic record series*(S) ≡ for all $F \in S$ and for all $X \in F$
Authentic record(X) $\wedge$ *Record series integrity*(S)

3. A Procedure for Ensuring the Integrity and Authenticity of Digital Record Series

The Java archive (JAR) format is a platform-independent file format that aggregates many files into one. JAR was developed so that Java applets and their components could be bundled into a single file (package) and quickly downloaded to a browser in an http transaction. The Java application launcher can launch one of the files, e.g., a class with a method, in the package. A JAR provides the capability to verify the origin of the components in the JAR so that only programs authored by persons or organizations trusted by the user will be executed. JAR is an open industry standard.[34]

In this section, an adaptation of the JAR file format to store digital records will be described. Fig. 1 shows an example of the directory structure of a JAR. The META-INF(ormation) directory contains three files. The META-INF directory is followed by the digital files corresponding to the digital records stored in the JAR.

```
META-INF/MANIFEST.MF
META-INF/SIGNATURE.SF
META-
INF/SIGNATURE.DSA
wp/corr/file1.wp5
wp/corr/file2.wp5
lotus/schedule.wks
lotus/budget.wks
photo/image1.jpg
photo/image2.gif
```

Figure 1. Directory Structure of package.jar

Fig. 2 shows an example of a manifest file MANIFEST.MF. The manifest file consists of a set of path names/file names for files and annotations of these files.[35] The annotations corresponding to a path/filename are called a section of the manifest. The message digests in the manifest file are created from the files themselves.

```
Manifest-Version: 1.0
Organization: Executive Office of the President
Organizational-Unit: Legislative Affairs, Office of
Name-of-record creator: "Richard Breeden"
Series-title: "Richard Breeden's Files"

Name: Chronological Correspondence/file1.wp5
SHA1-Digest: TD1GZt8G11dXY2p40lSZPc5Rj64=
File-format: wp5.1
Document-type: letter
Name-of-author: Breeden, Richard
Name-of-addressee: Kristol, W; Kolb, C
Chronological-date: 01/12/92
```

[34] Sun Microsystems, Inc. Java™ 2 SDK, Standard Edition Documentation, Version 1.3.1, 2001
[35] The manifest file of a JAR represents attributes and values in the form "header: value".

```
Archival-date: 01/12/92

Name: …
```

Figure 2. Manifest File

Fig. 3 shows the contents of the signature file (SIGNATURE.SF). The file includes the message digest for the entire manifest. It also contains digest values created from sections of the manifest file.

```
Signature-Version: 1.0
SHA1-Digest-manifest:
"hlyS+K9T7DyHtZrtl+LxvqgaMYM="
Created-By: Signature File JDK 1.2
Name: wp/corr/file1.wp5
SHA1-Digest: r58H40lDL39d6a2tU6T38Letz64=
```

Figure 3. Signature File

The file SIGNATURE.DSA is associated with the signature file with the same file name, but has a different file extension (DSA). This file stores the digital signature of the corresponding signature file and an X.509 certificate for the public key of the signature.

In the full technical report, a procedure (3.3) is described for storing received records in JARS, for retrieving records and for verifying their authenticity.

4. Proving the Correctness of Procedures

The theory of record authenticity described in section 2, when combined with theories of belief and communications security, can be used to prove the correctness of procedures for maintaining the authenticity and integrity of records. The method for proving the correctness of procedures is as follows:

1. Express the assumptions and goal of the procedure in the logical language.
2. Make assertions in the logical language as to what is true after the execution of each procedural step.
3. Apply the axioms, definitions and deduction rules to the assumptions and results of procedural steps to derive the goal, e.g., a record is authentic.

The following theorem establishes that Procedure 3.3 can be used to maintain and verify the authenticity and integrity of an digital record series archived in JARs.

Theorem 4.2: If a trusted record server for an organization receives from a principal of the same organization a message that contains a digital record, the record's classification code, and other statements about the record made by the principal, and the source and integrity of the message have been verified, and the principal is a trusted authority on the business activity in which the digital record was made or received, and the record server archives the digital record and other statements about the digital record using procedure 3.3, then in response to a request for a record from a member of the organization, the record server can use procedure 3.3 to verify the integrity and authenticity of the archived records.

Theorem 4.2 is proved by mathematical induction on the number of stored records. The rather lengthy proof of this theorem is in an appendix of the full technical report.

The significance of this and similar theorems is that their proofs demonstrate a formal method for analyzing communication, record-keeping and preservation procedures to determine whether they achieve integrity and authenticity goals. As often as not, the proof uncovers faults of the procedure, which upon repair may be proven correct.

5. Results and Research Issues

The concepts of record, record series and archival integrity, and of record and record series authenticity have been expressed as axioms or definitions in a logical language. This theory, when combined with theories of belief and communications security can be used to prove the correctness of procedures for maintaining the integrity and authenticity of electronic records.

Procedures for transferring records to the custody of an archival institution and preserving records in an archival system can be developed. These procedures can be formally analyzed to determine whether they also maintain the authenticity of the transmitted and preserved records.

Among the assumptions of Theorem 4.2 is that there are no preservation transformations on the records. If digital record format are migrated to other formats or file viewers are rewritten for new platforms, demonstration of record integrity will require demonstrating that these transformations preserve the content and essential elements of documentary form of the digital records.

The theory of documents and records presented in this paper is not complete, that is, every valid statement about digital records is not provable from the axioms. Also, the theory is probably not sound, that is, every theorem that is provable may not be valid. Furthermore, the independence of the axioms has not been demonstrated. A semantic model for the theory of records should be defined. Then an attempt should be made to modify the theory so that it is complete and sound. A formal semantics for the theory should also make it possible to provide a semantic analysis method for analyzing the authenticity property of record-keeping and archival preservation procedures.