

## Record Integrity and Authentication for Electronic R&D

Author: Michael H Elliott

*Note: The following is a summary of a comprehensive analysis on electronic record integrity and authentication contained in the publication **Electronic Laboratory Notebooks: A Foundation for Scientific Knowledge Management Edition III**. Further information can be obtained at [www.atriumresearch.com](http://www.atriumresearch.com)*

A patent provides rights to an inventor to restrict the ability of others to make, use, or sell an invention. In the scientific domains, laboratory notebooks are typically used as the primary evidence to prove inventorship of a concept and the details of its first successful use, or what is known as “reduction to practice.” The particulars of the research and associated dates and times are especially critical to establishing proprietary rights in the United States. The reason for this is that the U.S., versus other countries, awards patents on a “first-inventor” basis rather than on a “first-to-file.” This puts the responsibility in the hand of the inventor to have accurate and corroborated records to prove they created and successfully demonstrated the innovation before others. Being so critically relied upon, entries in a laboratory notebook must be clear to demonstrate how and when the work was performed, signed by the author, and corroborated by a witness not involved with the original work.

The use of Electronic Laboratory Notebook (“ELN”) technology has risen sharply in the past four years, having now penetrated over 20% of all biopharmaceutical companies.<sup>i</sup> ELN has not only led to increased laboratory efficiency and improved leverage of institutional knowledge, but also to enhanced protection of Intellectual Property (“IP”). Based on technology to manage the underlying electronic records, ELN *can* provide the access security, version control, record authentication, and automated time stamping their paper forbearers cannot.

In December 2006, the risk of ignoring the proper management of patent-supporting electronic records changed appreciably with amendments to the United States Federal Rules of Civil Procedure (“FRCP”). The FRCP changes alter the procedures of discovery, which is the process of requesting, or compelling, information from one party to another in a civil case. Since 97% of scientific records have their basis in electronic form, these changes effectively make all discovery now electronic discovery or “e-discovery.” This has wide-ranging implications which affects the retention of research data, IP records management practices, organization of data, and data storage formats.

To ensure the trustworthiness of electronic records, effective management controls must be in place to guarantee their timing, integrity, and authenticity. One must not only secure data against theft or alteration, but be able to prove when the record was created, who created it, who approved/signed it, detection of any changes, and an auditable trail of the record’s lifecycle. An organization must have consistent, audited, and proven record management practices that are enforced across the entire spectrum of their research operations.

The effective management and control of e-records is also important to assure their admission into a court proceeding. The records must pass a series of criteria to be classified as “business records” under the Federal Rules of Evidence. Not only do they have to be maintained under the normal course of the business, but they must be proven to be authentic to avoid being classified as “hearsay.” The US Federal Judicial Center’s *Manual for Complex Litigation*<sup>ii</sup> notes that a judge should “consider the accuracy of computerized evidence” and a “proponent of computerized evidence has the burden of laying a proper foundation by establishing its accuracy.” In the case *In Re Vee Vinhne*<sup>iii</sup> the appellate court affirmed the lower

court's denial of electronic records admission noting that the "focus is not on the circumstances of the creation of the record, but rather on the circumstances of the preservation of the record during the time it is in the file so as to assure that the document being proffered is the same as the document that was originally created." No longer can companies continue to permit scientists and technicians to arbitrarily manage records in an uncontrolled manner.

PKI digital signature and X9.95 timestamp technologies have come on the market in the last few years to help organizations minimize their risk exposures. In the following sections we will discuss these and other technologies employed for authentication.

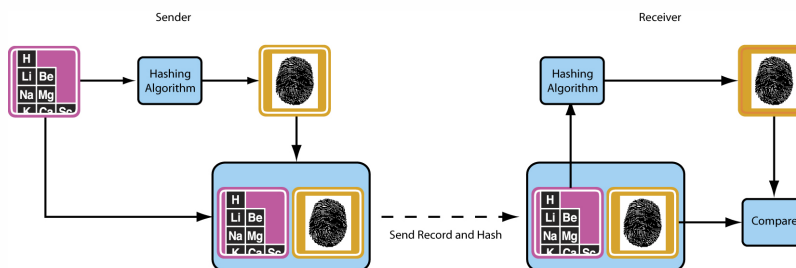
Record Authentication Technology

There are a number of different technologies used to establish the authenticity of records and users. There are electronic signatures, timestamps, hash digests, checksums, cyclic redundancy checks, and so forth. What technology is used and how it is deployed is dependent on the need of the particular user and the environment. In the scientific domains for patent protection and/regulatory compliance, electronic records must have the support of technology to prove that they are what they purport to be, any changes have been detected and logged, detection of corruption in transmission, and that they have a valid creator, sender, and receiver. Particularly with patents, the attestation of the time and date of the record's creation is crucial. In *Re Scott T. Jolley*,<sup>iv</sup> a patent was upheld principally on the basis of an e-mail timestamp.

A digital signature is "cryptographic process used to assure message originator authenticity, integrity, and non-repudiation."<sup>v</sup> Often confused with the broader "electronic signatures," digital signatures adhere to the principle of non-repudiation, which means there is no way for a person to deny that they created a record, they sent it, or they received it.

The concept of a *hash function* needs to be explained as this process is used natively and in combination with other signature processes. Cryptographic hash functions create an identifier based on the digestion of a record. Since this *hash value* is based on the contents of the record processed, it is unique. If the record content changes, the re-hashing of the record will result in a new value. The hash value is also known as a digital fingerprint due to this uniqueness.

**Figure 1 Hashing creates a unique fingerprint for a file**



There are many hash algorithms that have been used over the years such as MD5, RIPEMD-160, and SHA-1. RIPEMD-160 is most popular in Europe, while SHA-1 and SHA-2 are commonly used in the U.S.

Another important concept is encryption. Encryption is a cryptographic technique for obscuring a record to make it unreadable without special tools or knowledge to decrypt it. A *cipher* is used for the encryption and decryption process and in many cases a *key* is used to modify the cipher algorithm. Having the correct "key," or piece of code, will allow the algorithm to function properly. The analogy is a door lockset – the cipher is the lock itself, while the key either allows or restricts the ability for the door to open.

In *symmetric* encryption, or “secret key” cryptography, the same key is used for encrypting and decrypting the record. In other words, you hand the key over to the person on the other side of the door to unlock it. In *asymmetric* encryption, or “public-key” cryptography, different keys are used. In the case of a digital signature, a *private key* is used to encrypt the file and a *public key* is used to decrypt it. These keys adhere to the concept of non-repudiation; each key cannot undo its own particular function. Once a file is encrypted, the private key cannot unlock it. In the door analogy, your key to lock the door is different than another’s key on the other side used to unlock it.

*Public Key Infrastructure*

The Public Key Infrastructure (“PKI”) X.509 standard uses digital signatures based on asymmetric encryption. The goals of PKI are to create a trusted relationship between one party and another to authenticate the parties, guarantee the integrity of a data transmission, and to ensure data privacy.

PKI utilizes the concept of “digital certificates” which are a form of credit card or passport which identifies certain characteristics of the sender. These certificates act as a form of guarantee to prove the authenticity of the sender. A user’s (or subject data) identity is matched to a public key and the details about the encryption algorithm are contained in the certificate. These are issued by a “Certificate Authority” (“CA”) which verifies the identity of user and include the digital signature and other information of the authorizing agent. Just like a credit card, these certificates have a specified period of validity and can be revoked. The exact implementation of PKI is a bit unique to the specific provider of the technology.

Though very robust and well established in the market, the use of PKI is not without its challenges. The infrastructure to support public key cryptography can be quite daunting for small or medium size organizations. Not only do the costs of the technology have to be considered, but the policies, procedures, and administration have to be taken into account for the total cost of ownership. It’s a matter of risk analysis and a balance between costs and the potential exposure to your data.

Certificates by definition have a limited life; if they didn’t, they wouldn’t be very useful. A significant advantage of the technology is that certificates are revoked if someone leaves the company, loses their private key, or misuses the system. As shown below, during the period of validity, a user who signed a document, like an ELN page, can be easily verified. But after they are revoked or expire, the certificate authority cannot truly verify the user.

**Figure 2: PKI certificates have a defined lifespan**

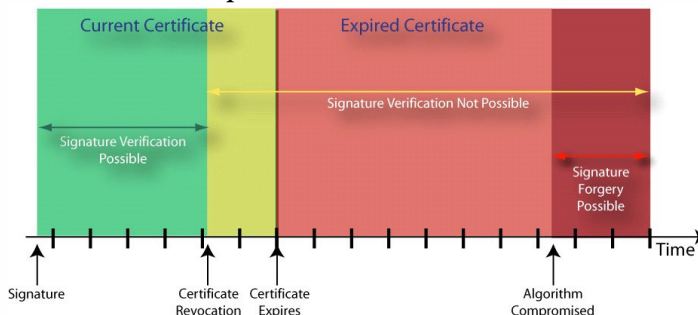
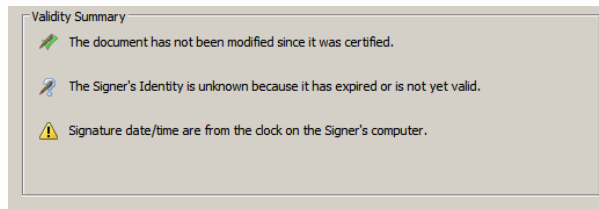


Figure 3 shows a how a PDF treats an expired certificate. In the second line you will note the comment “Signer’s Identity is unknown because it has expired or is not yet valid.”

Figure 3: Signature properties from Acrobat for a PDF with an expired certificate for a signer



Over time, algorithms can be compromised, just like the MD5<sup>vi</sup> and SHA-1<sup>vii</sup> hash algorithms have been. According to Landon Curt Noll of SystemsExperts Corporation, “By far, the services that are most vulnerable to the recent attacks (on MD5 & SHA1) are digital signatures and related document authenticity signatures.” It is also possible, in the absence of adequate security management practices, that private keys could be lost or stolen and their passwords hacked. CAs could be tampered with, either via an internal or external hack, creating false certificates. Without the use of a trusted time authority, dates and times of records could be back dated without anyone’s knowledge.

#### Signatures and Authentication For Everyone (“SAFE”)

An exciting development for digital signature standardization is the “Signatures and Authentication For Everyone” initiative, or SAFE. SAFE endeavors to streamline digital authentication and rights management primarily in the biopharmaceutical industry. Faced with a complex matrix of overlapping and potentially conflicting digital signature products, SAFE unifies authentication for system sign-on and record digital signature.

SAFE is PKI-based, compatible with the X.509 standard. SAFE accredits selected PKI Certificate Authorities that meet the established SAFE standard for credential services. In this manner, CAs can exist behind the firewall of an organization or be hosted by a third party. Any accredited organization using a certified authority can establish secure data transmittal with another using a Safe Bridge Certificate Authority (“SBCA”). SAFE requires the use of a hardware identity device such as a Smart card or USB token for a private key linked to a specific individual.

Though extremely beneficial to the industry, the SAFE standard does not relinquish the requirements for sound electronic record management practices. SAFE primarily provides an infrastructure for the authentication of identities. It can also be used, via PKI and its digital fingerprint, to determine if a file has been altered once it has been signed. However, for intellectual property protection, records must be proven to be trustworthy and authentic from point of creation. As noted earlier, certificates also expire over time, making it difficult to properly identify the signatory after certificate expiration.

#### Time Authentication

The time of record birth may be crucial in the defense of a patent, particularly in the U.S. The concern with PKI is that the date/time synchronization of the Certificate Authority is by policy only. In reality, there is nothing in PKI to prevent backdating of server time. There is no requirement a PKI enabled application must also get its timestamp from a trusted source. The application could, in theory, get the time of record creation from the client clock which could be capriciously set by a user.

#### *RFC 3161*

The Internet Engineering Task Force (“IETF”) proposed RFC 3161 Internet X.509 Public Key Infrastructure Time-Stamp Protocol (“TSP”) standard was created to resolve the conflicting methods of time attestation used in digital signatures. An



expansion to the X.509 PKI standard, RFC 3161 specifies the use of a Time Stamp Authority (“TSA”) and “timestamps.” A TSA is a trusted third party or an internal time source synchronized with a trusted time source that provides timestamp “tokens,” which are generated via a message and hashed time combination. This creates non-repudiable evidence that a record existed before the time of the token creation. The token is then appended to an X.509 digital signature.

This process provides a timestamp *near the time of the signature* but not a non-repudiable date of creation for the actual record itself. The date of invention is therefore based on some arbitrary date *when the record was signed, not the true date of when the work was performed*. Other records which might support IP (e.g., mass spectrum, chromatography, gene sequences, e-mails, etc.) often do not undergo a signature process in non-GxP regulated laboratories. These records, which can fall under the changes in FRCP and the Federal Rules of Evidence, may have to be produced in court. If these are being managed randomly or differently than your ELN records, this can raise doubts about your overall level of data management integrity to a judge or jury. Considering that the *In Re Jolley* turned on an e-mail to prove the date of invention, it is not a simple matter of only notebook pages requiring accurate and authenticated dates and times.

#### X9.95

Acknowledging the limitations in X.509, the financial services industry - which is highly dependent on recording time - has specified the X9.95 “Trusted Timestamp” standard under the governance of the American National Standards Institute (“ANSI”). X9.95 was built upon RFC 3161 and the International Standards Organization (“ISO”) 18014 time stamp standards. It differs from RFC 3161 in that it is independent from a public key infrastructure, leading to increase portability of the timestamp. In documents describing the standard, ANSI says in reference to PKI:

“...neither the symmetric nor the asymmetric cryptographic mechanisms described provide timeliness. The timing of the data generating events, such as the generation of the digital data or the generation of the digital signature itself, cannot be verified. If a timestamp is included as part of the digital data, some evidence is provided as to when the data generating event occurred. If, however, the clock used to generate the timestamp is under the control of the data event generator, the timestamp and therefore its timeliness is easily subject to manipulation and is therefore suspect. Only a timestamp token can provide timeliness that is verifiable and provable to a third party.”<sup>viii</sup>

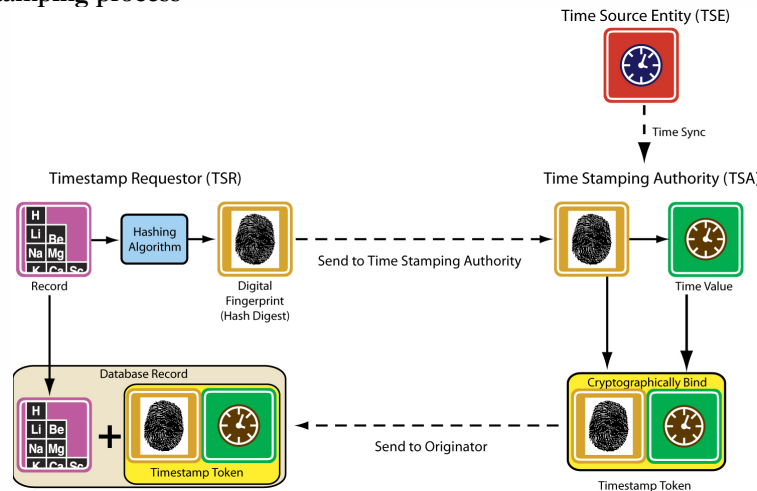
X9.95 defines the complete package of management practices, policies, responsibilities, and technical requirements for timestamp non-repudiation and authenticity. The X9.95 process is shown To validate if a record or time value have been altered, the process is reversed. First, the timestamp token is split into the hash digest and time value. Next, the record is hashed, creating another hash value. This value is compared to the hash from the timestamp token. If they compare favorably, the record has not been tampered with. If they do not, the record is considered altered and invalid.

Figure 4. First, a hash digest of the electronic record is created using algorithms like RIPEMD-160 or SHA-2. This fingerprint is sent to the Time Stamp Authority. The TSA generates a time value for the hash digest and cryptographically binds the time value and the hash together creating a “timestamp token.” This is sent back to the originator and linked to the original record. This process can be used to create timestamp tokens for any records in an intellectual property hierarchy, like spreadsheets, e-mails, instrument data files, or other records which are not digitally signed. One of the challenges of this method, though, is to keep up to date with the progression of hash standards and to maintain backward compatibility.

To validate if a record or time value have been altered, the process is reversed. First, the timestamp token is split into the hash digest and time value. Next, the record is hashed, creating another hash value. This value is compared to the hash

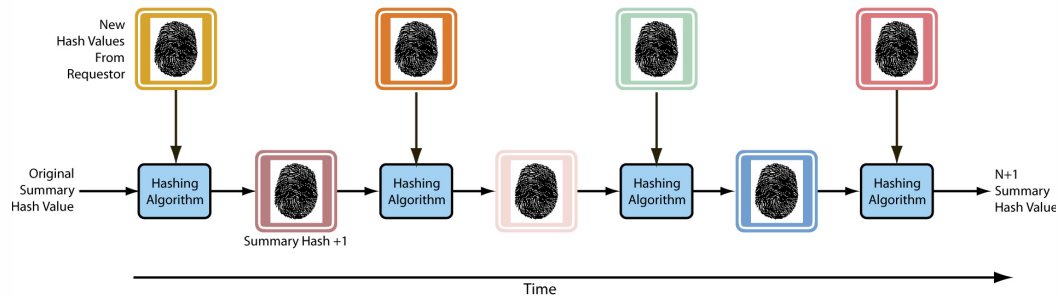
from the timestamp token. If they compare favorably, the record has not been tampered with. If they do not, the record is considered altered and invalid.

**Figure 4: Basic time stamping process**



However, Stuart Haber and Wakefield Scott Stornetta proved the vulnerability of the basic time stamping process. They noted that if the TSA deceptively modified the time of a particular token, the change would go undetected.<sup>ix</sup> One of the approaches used to address traceability of tokens is the “linked token” method. This is where the hash digest from the requestor is linked to all those generated in the past at the TSA. A “summary hash” which is the digest of all hashes up to that point, is hashed with the message digest, creating a new summary value. As shown below, this process repeats infinitum. The advantage of this approach is traceability, as it creates a non-repudiable chain of all values to date. If any of the underlying hash values changes in anyway, the summary hash – to whatever point in time is selected – would not match. This summary hash is stored with, or in, each timestamp token.

**Figure 5: Linked token method**



The linked token method is built from the concept of Merkle Trees. In a Merkle, or “Hash” Tree, hash pairs are hashed; those resulting hashes are paired up and hashed; and this process continues up a “tree” ultimately resulting in a single “root” hash value. If any of the underlying files or hash values change – or their sequence is altered - the root hash would be different than the one previously generated. This root hash is generally published on a regular basis in a public space like a newspaper.

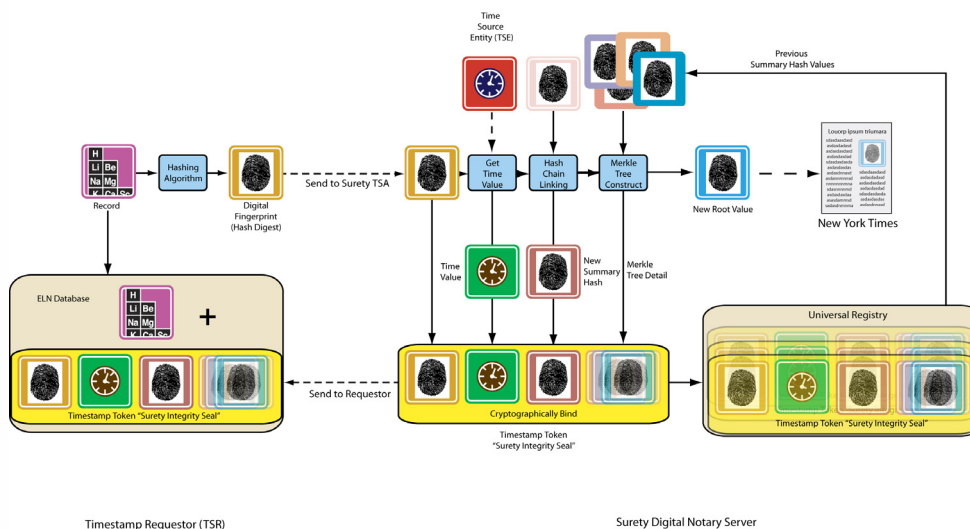
## Surety LLC

The X9.95 and ISO 18014-3 compliant third-party timestamp service company that has been most active in the space of intellectual property protection and ELN for the past several years has been Surety LLC based out of Herndon, Virginia. IP.com, a third party provider of IP search, verification, and storage solutions uses the Surety AbsoluteProof technology for trusted time-stamping and notarization. ELN suppliers EKM, KineMatik, reDesignLive, and Symyx as well as several electronic content management and active archive vendors have integrated with the Surety process; they either OEM this service into their application or provide it as an option to their customers.

Surety utilizes both the concept of linked tokens and Merkle Trees for non-repudiation. Different from basic timestamp authorities, Surety stores each record hash value, the timestamp token, and summary hash value into what they call the “Universal Registry” which is a database located at the company. Illustrated in Figure 6, a requestor (e.g., an ELN, SDMS, or LIMS) sends a hash digital fingerprint to the TSA. Just like any other TSA, the time value is cryptographically bound to the hash digest, creating the timestamp token. However, Surety also performs hash chain linking, creating a new summary value and storing it in the token. The summary hash from the linking and the hash values that made up the tree are included in the timestamp token, creating what Surety calls an Integrity Seal (“Seal”).

Weekly, the summary hash value is published in the public or commercial notices section of the *New York Times*, providing the opportunity for an independent audit of the Universal Registry’s integrity. Since all the original and summary values are stored, the root can be reconstructed upon demand. If the reconstruction of the tree root matches the value published in the *Times*, then the integrity of the TSA is undeniable, addressing the anxieties of Haber and Stornetta.

**Figure 6: Surety notary process**



Surety offers a freely available program to verify the integrity of a token. To verify, the requestor sends the token (a.k.a. “Seal”) back to the TSA (a.k.a. “Digital Notary Server”). On the Surety side, the server re-computes the root hash value. It then combines the root hash value with the previous hash values in the Seal. The actual value is retrieved from the Universal Registry and compare to the calculated value. If they match, then the file is considered trustworthy.

The resistance to “collision” is the major consideration for hash algorithms. Collision means that the same hash digest can be generated for the same file. Since hackers are continuing in their efforts to crack these algorithms, academic and government agencies are working diligently to test their robustness. Standards groups like NIST recommend the use of newer, stronger algorithms like the more complex SHA-2 over the older SHA-1 because of the lack of collision resistance. To maintain backward compatibility, Surety has patented a process for creating new Integrity Seals as these standards evolve. Upon demand, a user can “Renew” their timestamp tokens using the new hash method which generates a new digest, summary hash, and root value without affecting the timestamp.

### Summary

In summary, electronic records are increasingly being used in court proceedings throughout the world. The December 2006 changes to the US Federal Rules of Civil Procedure now explicitly address the legal discovery of electronic records. These changes affect all organizations who file patents in the US. Electronic records must also be proven to be authentic and accurate under the Federal Rules of Evidence.

Two complementary methods of record authentication are increasingly being used by biopharmaceutical research organizations in conjunction with electronic laboratory notebooks. SAFE is a newer development to standardize PKI X.509 digital signatures throughout the pharmaceutical industry. However, SAFE does not provide long term assurance of the date and time of record creation, nor does it apply broadly to most scientific records which do not go through a digital signature process. The ANSI X9.95 standard is a complementary process to PKI, providing a non-refutable timestamp on any electronic record. In our view, risks are reduced by the proper use and maintenance of both technologies. However, this does not diminish the critical need for organizations to have sound and proven record management policies, record retention procedures, and audited and compliant business practices.

As Chief U.S. Magistrate Judge Paul Grimm (US District Court Maryland) wrote in his opinion in *Lorraine et al v. Markel American Insurance Company*<sup>x</sup>, “Further, although ‘it may be better to be lucky than good,’ as the saying goes, counsel would be wise not to test their luck unnecessarily. If it is critical to the success of your case to admit into evidence computer stored records, it would be prudent to plan to authenticate the record by the most rigorous standard that may be applied. If less is required, then luck was with you.” If you aren’t ready to prove the authenticity of your electronic records, you’re increasing your risk exposure - daily.

***NOTE: This material is ©2007 Atrium Research & Consulting. This material was licensed with permission. Its use in no way indicates an endorsement of Surety LLC products and/or strategies by Atrium Research. Trademarks are the property of their respective owners.***

<sup>i</sup> Elliott, Michael, *2006 Electronic Laboratory Notebook Survey*, Atrium Research & Consulting, Wilton CT USA

<sup>ii</sup> Federal Judicial Center, *Manual for Complex Litigation Fourth Edition*, 2004 Washington D.C.

<sup>iii</sup> *In Re Vee Vinbee*, 336 B.R. 437 (9<sup>th</sup> Cir. BAP 2005) Lexis 2602

<sup>iv</sup> *In Re Scott T. Jolley*, 308 F.3d 1317, 64 USPQ.2d (BNA) 1901 (Fed. Cir. Oct. 29, 2002) (Interference Nos. 103525 and 103526) <http://www.ll.georgetown.edu/federal/judicial/fed/opinions/01-opinions/01-1646.html>

<sup>v</sup> National Information Systems Security (INFOSEC) Glossary, NSTISSI No. 4009, January 1999 (Revision 1)

<sup>vi</sup> Kaminsky, Dan, *MD5 to be Considered Harmful Someday*, December 2004 [http://www.doxpara.com/md5\\_someday.pdf](http://www.doxpara.com/md5_someday.pdf)

<sup>vii</sup> Noll, Landon Curt, *SHA1 Cryptographic Hash Update*, SystemsExperts Corporation December 2005 <http://www.systemexperts.com/tutors/CryptographicHashUpdate.pdf>

<sup>viii</sup> *American National Standard for Financial Services Working Draft X9.95-040304*, <http://www.oasis-open.org/archives/dss/200408/doc00000.doc>

<sup>ix</sup> Haber, Stuart and Wakefield Scott Stornetta, *How to Time Stamp a Digital Document*, Journal of Cryptology 1991 3 (2)

<sup>x</sup> *Lorraine et al v. Markel American Insurance Company*, 1:2006cv01893 (US District Court Maryland 2006)