Pour publication dans Annales des Télécommunications, Mai/Juin 2006.

The digital signature dilemma

Le dilemme de la signature numérique

Jean-François Blanchette[*]

Abstract

The last ten years have seen an enormous amount of legal, regulatory, and technological activity aimed at designing a proper electronic equivalent to handwritten signatures. One such design, that of cryptology-based (or digital) signatures, has succeeded over other solutions to the point where, in certain legal systems, such as those of the Member States of the European Union, electronic signatures are almost exclusively understood to be based on public-key cryptography. Yet, several archival institutions (including the National Archives of Canada, Australia and the US) have expressed ambivalence at the prospect of preserving digitally signed records. This paper argues that discrepancies between technical, legal and archival responses to the problem of long-term preservation of digitally signed documents are founded on diverging understandings — *physical* vs. *contextual* — of electronic authenticity.

Keywords: digital signatures, evidence law, electronic records, archivage, cryptography.

Résumé

Depuis dix ans, d'énormes efforts ont étés déployés sur le plan juridique, technologique et législatif dans le but d'élaborer un équivalent électronique à la signature manuscrite. Un des mécanismes proposé à cet effet est celui de la 'signature numérique', fondé sur les technologies de cryptographie à clé publique. Dans certain systèmes juridiques (p.ex., ceux des États Membres de l'Union Européenne), l'approche cryptographique a rencontré un tel succès auprès des législateurs que la signature électronique s'y comprend presque exclusivement en termes de cette méthode. Néanmoins, plusieurs institutions archivistiques (entre autres, les Archives Nationales du Canada, de l'Australie et des Etats Unis) ont exprimé une certaine ambivalence à l'idée de préserver des documents d'archives signés numériquement. Cet article propose que les différences entre les propositions techniques, juridiques, et archivistiques face au problème de la préservation de documents numériques signés sont fondées sur des conceptions divergentes de l'authenticité électronique — *physique* versus *contextuelle.*

Mots clés: signatures électroniques, droit de la preuve, documents électroniques, archivage, cryptographie.

[*] Department of Information Studies, University of California, Los Angeles, Box 951520, Los Angeles, CA 90095-1520, USA; Email: blanchette@ucla.edu; Web:http://polaris.gseis.ucla.edu/blanchette.

I. Introduction

The very fluidity that makes e-commerce potentially so enormous, its ability to seamlessly cross over borders and traditional market boundaries, is also its greatest liability: How can parties establish trustworthy relationships in shifting environments, characterized by the absence of traditional methods for establishing identity, commitment, evidence, and trust? In the world of paper-and-ink contracts, these objectives are typically achieved through the use of a most mundane technology, handwritten signatures.

A primary purpose of signatures, be they traditional, handwritten, ones, or based on esoteric mathematical algorithms, is to serve as instruments of law, as the preferred instrument for parties to manifest their consent and provide proof of their respective commitments. Signing is, of course, within most legal texts, understood to be concomitant with the use of paper as the instrumentum, the physical means whereby contractual agreements are inscribed, preserved, and, most importantly, exhibited during disputes. The last 10 years have thus seen an enormous amount of legal, technological, and legislative activity aimed at designing a proper *electronic equivalent* to handwritten signatures. One such design, that of cryptology-based electronic signatures,[1] has succeeded over other solutions to the point where, in certain legal systems, such as those of the Member States of the European Union, electronic signatures are almost exclusively understood to be inevitably "digital signatures", that is, based on cryptological solutions, more specifically, public-key (or asymmetric) cryptography [1, 2].

However, the efforts of the legal and technological community at enshrining digital signatures as the exclusive substitute for handwritten signatures has not met with its expected success on at least two fronts: on the one hand, predicted markets for digital signature technologies and public-key infrastructures have largely failed to materialize [3]; on the other hand, the archival community, the very community historically entrusted with the care and preservation of documentary evidence, has developed intellectual tools and practices which supports an understanding of electronic documentary evidence as primarily *contextual*, rather than the primarily *physical* understanding supported by digital signatures. This paper reviews the evolution of these two divergent notions of electronic documentary evidence as they have been expressed through various laws, taking both of these understandings into account.

Section II reviews digital signature technology, and the model it proposes for an electronic equivalent to handwritten signatures; section III reviews how this model was transposed into evidence law in the EU and in the US; section IV discusses the electronic signature lifecycle, raising three technical issues entailed by the problem of preserving digitally signed documents over time; section V reviews the various solutions offered by the technical community to resolve these issues, i.e., "trusted archival services", "resignature", and "canonicalization"; section VI discusses the responses of the archival community to those same issues, reviewing documents from the US National Archives and Records Administration, Library and Archives of Canada and the National Archives of Australia, as well as the conclusions of a research project founded in the archival world, InterPARES; section VII concludes with some reflections

[1] The established (if confusing) terminological usage is that "digital signatures" refer exclusively to those based on public-key cryptography, while "electronic signatures" refer to all potential technologies, including biometrics, etc.

on the road ahead.

II. Digital Signature Technology

Up until thirty years ago, cryptology essentially remained a military science, providing technologies to generals, diplomats, and spies wishing to communicate privately. In the 1960s, the security needs of the banking industry spurred the emergence of an academic cryptology research community, independent from the intelligence establishment [3, 4, 5, 6]. In 1976, this community made its presence widely known, with the publication of Diffie and Hellman's "New Directions in Cryptography."

In this seminal paper, the authors introduced not only a radically new method of key exchange, but also the concept of public-key cryptography, widely acknowledged as one of the most important development of modern cryptography, and finally, suggested how public-key cryptography could be used to offer not only *confidentiality*, but also, *authentication* services: "in order to have a purely digital replacement for [written contracts], each user must be able to produce a message whose authenticity can be checked by anyone, but which could not have been produced by anyone else, even the recipient" [7].

In a nutshell, public-key cryptography functions by assigning two keys (private, public) to every user on a computer network: the *private key* can only be legitimately accessed by its owner, while the *public key* is made available to other users on the network through publicly accessible directories. The unique advantage of public-key cryptography rests on the fact that while the private and public keys are mathematically related, *knowing the public key, it is computationally infeasible to deduce the private key.*[2] Such a system can be used to perform two distinct tasks: (a) encryption and (b) authentication:

1. To transmit a *confidential* electronic message over the network to user Bob, user Alice encrypts the message using Bob's public key, before sending it to him. Because of it's unique mathematical relationship to its public counterpart, only Bob's private key will successfully decrypt the message;

2. To "*sign*" a message, the role of each key is reversed: Alice encrypts the message using her private key before sending it to Bob. If Alice's public key successfully decrypts the message, Bob is then be convinced that only Alice (or rather, Alice's private key) could have signed that message. If the decryption fails, either the message was not signed using Alice's private key, or the document was modified — even by a single bit — at some time after the signature was created.

The cryptological model for digital signatures is thus characterized by a signing algorithm, requiring the signer's private key, and a verification algorithm, requiring the signer's public

[2] Modern cryptography makes extensive use of a number of *computational assumptions*, that is, hypotheses regarding the difficulty of solving certain mathematical problems (e.g., determining the prime factors of large numbers). While it is not known for certain whether these problems are genuinely 'difficult', no one has yet claimed to have found a solution for them, and thus, solving them is deemed to be "computationally infeasible."

key. Because the signer's public key is openly available on the network, users need not communicate prior to exchanging signed messages, thus providing an efficient system for securing commercial transactions. In practice, the use of digital signatures within organizations requires the deployment of public-key infrastructures (PKI), the enabling ensemble of software, hardware and procedures providing the necessary key management, directory and revocation services.[3]

III. Digital Signatures and Evidence Law

Widespread acceptance of the cryptological model of electronic signatures could only have occurred based on a number of factors: (1) legal texts which specifically required that *written* signatures be used in transactions had to be modified; (2) the strict controls regulating the use of cryptological technologies had to be softened, or altogether abandoned. Given the nature of the institutions in play (law, intelligence agencies), such changes should have taken decades to achieve, but the mid-nineties explosion of the Internet on the world scene, and the ensuing e-commerce "tidal wave" insured that, all over the world, governments lent a much readier ear to calls for adapting their legislations and softening up cryptology control laws, in order to ensure the most favorable environment for the blossoming of e-commerce.[4] Very different approaches to this complex adaptation gradually emerged at the international level, in the United States, and in the European Union.

III.1. UNCITRAL Model Law on E-commerce

The United Nations Commission on Trade Law (UNCITRAL) is a UN organization with headquarters in Vienna. Created in 1966, the UNCITRAL is composed of thirty-six member States elected by the General Assembly, representative of the world's various geographic regions and its principal economic and legal systems. The UNCITRAL Model Law on electronic commerce was adopted in 1996, with the objectives of "facilitat[ing] the use of modern means of communications and storage of information, such as electronic data interchange (EDI), electronic mail and telecopy, with or without the use of such support as the Internet. It is based on the establishment of a functional equivalent for paper-based concepts such as 'writing', 'signature' and 'original.' By providing standards by which the legal value of electronic messages can be assessed, the Model Law should play a significative role in enhancing the use of paperless communication."[5]

The most fundamental principle of the Model Law is that of "non-discrimination": Article 5 of the Model Law states that "[i]nformation shall not be denied legal effect, validity or enforce-ability solely on the grounds that it is in the form of a data message." The Model Law offers a *functional* definition for signatures, stating that "the signing method must enable one to identity the signer, and indicate that the signer manifests his consent." The Model Law has

[3] See [8] for a full description of the necessary elements of a public-key infrastructure.

[4] See [9] for an international review of the deregulation process of cryptographic technologies.

[5] UNCITRAL Model Law on Electronic Commerce with Guide to Enactment, A/CN.9/WG.IV/WP.88, November 1996.

been a very influential document, cited as a reference by most electronic signature legislations and the principles of "non-discrimination" and of a "functional" definition of signatures have enjoyed widespread dissemination, as effective legal devices to negotiate the transition between the requirements of the paper-and-ink world, and the promises of the new electronic worlds.

III.2. E-Sign and UETA

In the United States, the American Bar Association took an early lead in addressing the issue of electronic signature legislation, by publishing "Digital Signature Guidelines" [10]advocating the recognition of digital signature as the only valid form of electronic signature. The guidelines were adapted by the Utah legislature which became the first US legislature to adopt electronic signature legislation in 1995.[6]

Hoping to foster uniformity in this rapidly evolving area of legislation, the National Conference of Commissioners for Uniform State Law (NCCUSL) drafted in 1999 the "Uniform Electronic Transaction Act" (UETA), with the expectation that it would be adopted by all 50 states. Some, like California, did adopt it, but only after modifications so significant as to negate the desired harmonization.[7]

The "Electronic Signature in Global and National Commerce Act" [8] (E-Sign), enacted by President Clinton in 2000, sought to enforce a uniform legal framework for electronic transactions in the United States. E-Sign, just as UETA, adopted a broad definition of electronic signature, as "an electronic sound, symbol, or process attached to or logically associated with a record and executed or adopted by a person with the intent to sign the record." In order to enforce a technologically-neutral approach, states that passed technology-specific legislation (such as Utah) would see their legislation be pre-empted by E-Sign.[9] The pre-emption takes effect *unless* a state has adopted the UETA, in which case the UETA is applicable [11]. As of December 2004, forty-six states and the District of Columbia have adopted the UETA, with the remaining four (4) states (Georgia, Illinois, New York, and Washington) having enacted their own electronic signature laws. Broadly speaking, UETA has thus become the predominant law of the land with regard to electronic signatures.

[6] Utah Digital Signature Act, Utah Code Ann. 46-3-101 to 602 (2004).

[7] Cal. Civ. Code 1633.1 to 1633.17

[8] 15 U.S.C. 7001-7031 (2004).

[9] "A State statute, regulation, or other rule of law may modify, limit, or supersede the provisions of [E-Sign] with respect to State law only if such statute, regulation, or rule of law … specifies the alternative procedures or requirements for the use or acceptance (or both) of electronic records or electronic signatures to establish the legal effect, validity, or enforceability of contracts or other records, if … such alternative procedures or requirements do not require, or accord greater legal status or effect to, the implementation or application of a specific technology or technical specification for performing the functions of creating, storing, generating, receiving, communicating, or authenticating electronic records or electronic signatures." *ibid.*, sec. 102.

III.3. European Union Directive

The EU has adopted on December 13, 1999 "a European Parliament and Council directive on a community framework for electronic signatures."[10] Given the transnational potential of electronic commerce, the European Parliament sought to rapidly establish a harmonized legal framework and avoid any obstacles to the promised expansion of the European Internal Market. At the same time, European regulators hoped to repeat the economic miracle of the GSM cellular telephony standard and provide a regulatory framework which could kick-start the nascent market for electronic signature products and related services.

In order to achieve this dual objective, the Directive defines two distinct kinds of signatures:

- *Simple electronic signatures* are defined as "data in electronic form which are attached to or logically associated with other electronic data and which serve as method of authentication;"[11]
- *Advanced electronic signatures* "means an electronic signature which meets the following requirements: (a) it is uniquely linked to the signatory; (b) it is capable of identifying the signatory; (c) it is created using means that the signatory can maintain under his sole control; (d) it is linked to the data to which it relates in such a manner that any subsequent change of the data is detectable."[12]

While the first definition allows for a wide range of technologies, the second one is clearly directed at cryptographic signatures, since it is the only kind that fulfills mandate (d).[13] To create an incentive for market adoption of cryptographic signatures, each type of signature is granted a distinct evidential value: simple electronic signatures must be admitted in court, but the Directive does not specify their proof value; advanced electronic signatures must not only be admitted as evidence, but Member States must grant them an evidential value equivalent to that previously accorded to handwritten signatures.[14]

Since the mid-nineties, dozens of countries around the world have amended their evidence law in order to account for electronic signatures, with a significant number adopting regulatory schemes along the lines of the European Directive. Even in countries which have opted for more technologically-neutral approaches to evidence law reform, such as the United States, digital signatures and PKI have been offered as the technological foundation for the provision of online governmental services (see section VI below). The next section analyses the implications of cryptographic signature technologies for electronic document preservation.

[10] EC directive 1999/93/ec of the European Parliament and Council on a community framework for electronic signatures, Official Journal of the European Communities L 13/12 19. 1. 2000.

[11] *Ibid.*, Art. 2.1.

[12] *Ibid.*, Art. 2.2.

[13] That is, signature verification will fail if the signed document is modified — even by a single bit — after the signature is applied.

[14] *Ibid.*, Art. 5.

IV. The Electronic Signature Lifecycle

Governmental administrations, businesses, and individuals are obligated to preserve the records which prove their rights and define their obligations, so that they may be used as evidence if and when, at a later time, disputes arise over transactions. For example, the "Federal Records Act" mandates every US federal agency to "make and preserve records containing adequate and proper documentation of the organization, functions, policies, decisions, procedures, and essential transactions of the agency and designed … to protect the legal and financial rights of the Government and of persons directly affected by the agency's activities."[15] Such records may consist of, among other things, letters, receipts, contracts, memorandums, or in fact, any "data or information *in a fixed form* that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference."[16]

Given that the ability for records to serve as evidence hinges on this crucial characteristic of fixity, their preservation involves protection against two different kinds of threats: (a) natural decay and (b) intentional attempts to modify the information on records. In the case of paper, such protection involves well-know parameters: using adequate media and ink (protection against material decay), some form of cataloguing and indexation (protection against decay of institutional memory), access control (protection against malicious modifications), and the use of experts to ascertain the integrity of questioned documents.

In the case of signed electronic documents, the parameters are somewhat different, and our experience with such protection is much more limited. Two main differences with the world of paper documents are that (a) material preservation implies protection against both *media decay* and *format obsolescence*, that is, the magnetic or optical media underlying the electronic documents must be periodically renewed, and the encoding formats migrated, in order to ensure that documents may still be read, despite hardware and software obsolescence; and (b) the evidence created by the electronic signature must be preserved along with the document itself. In the case of cryptographic signatures, this implies that the preservation of all of the elements necessary for the process of signature verification.

These differences are made more explicit by looking at the lifecycle of a cryptographic signature, which can be broken into four distinct steps:

1. *Creation:* the cryptographic signature is created by the signer; the signed document is then sent to the person meant to receive it;

2. *Initial verification:* upon receiving the electronically signed document, the destinatory verifies the signature, and if successful, proceeds with the actions related to the document;

[15] 44 USC Ch. 15, § 3101.

[16] Richard Pearce-Moses, *A Glossary of Archival and Records Terminology*, Society of American Archivists, 2005 (emphasis added). A comprehensive definition of what constitutes a record is still a matter of debate within the archival community — see [33].

3. *Archiving:* the document and its signature are both archived with view of preserving them as evidence in potential future litigation;

4. *Litigation:* litigation does occur, the document is presented as evidence in front of a judge, and the signature verified again, so that the identity of the signer and the integrity of the document ascertained.

Of course, while step 4 may only occur rarely, if at all, the entire point of the archiving process (apart from questions of institutional memory) is to provide for just such an event. A number of important problems arise because of the significant time that may elapse between step 2 and step 4. That is, while the initial verification may occur within seconds, minutes, or days of the signature creation, the later verification will occur potentially years after signature creation, and in the context of an archived document. In terms of the evidence provided by a cryptographic signature, three distinct implications may be distinguished:

1. *Decay of security:* as a consequence of scientific advances in cryptanalysis, the initial cryptographic keys used for signature may become, over time, vulnerable, and thus enable forgery of signatures;

2. *Availability of verification software:* compatible software for signature verification must remain available over the entire lifetime of the document;

3. *Interaction between signature verification and document preservation*: cryptographic signatures freeze the signed document in its original state, forbidding any modification to its *bitwise integrity.*

This last implication is particular significant for the archival profession. Current practice for ensuring the intelligibility of electronic documents over time proceeds through updating their logical format (i.e., migration), so that they remain compatible with available software and hardware necessary to decode and render them on screen or on paper. Such migration will necessarily invalidate the signatures affixed to the documents, as the verification algorithm makes no difference between modifications resulting from an archivist, or from dishonest parties.

Herein lies the archivist's dilemma: ensuring two technologically incompatible missions, preserving the readability of documents, or that of the digital signatures affixed to them. As the next section details, these issues have received uneven consideration from the technical community.

V. Technical Responses

While the consequences of security decay due to advances in cryptanalysis have been extensively commented upon (for example, [12]) the fundamental dilemma facing archivists seeking to preserve the legibility of both documents and their cryptographic signatures has largely failed to surface in the technical literature. The technical responses offered to solve the problem of ensuring the long-term preservation of digitally signed documents have (so far) fallen under three distinct headings: (1) so-called "trusted archival services", (2) resignature, and (3) canonicalization.

V.1. Trusted Archival Services

The concept of "Trusted Archival Services" (TAS) was introduced in the context of the "European Electronic Signature Standardization Initiative" (EESSI) consortium, a standardization effort which seeks to translate the requirements of the European Directive on electronic signatures into European standards [13, 14]. The concept refers to a new type of commercial service that would be offered by emergent bodies and professions,[17] in order to guarantee the long-term integrity of cryptographically signed documents.

One EESSI report lists several technical requirements which such archival services would be expected to meet, among them, "backward compatibility" with computer hardware and software, through either preservation of equipment and/or emulation: "Trusted Archival Services (TAS) should maintain a set of applications (viewers as well as signature validation applications) together with the corresponding platforms (hardware, operating systems, etc) or at least an emulator of such applications and/or platforms *in order to guarantee that the content of the documents can still be viewed and that the signature on these documents can still be validated years later (even if the technology is not available anymore at that time)*."[18] (emphasis added)

Thus, the EESSI reports proposes that in order to the solve the problem of simultaneous preservation of documents together with their signatures, TAS act as information technology museums or invest in emulation strategies. This is because, as described in section IV the simplest and most widely accepted archival strategy, that of logical encoding migration, is not available for digitally signed documents.

No archival institution is seriously considering using original software and hardware, either through their *preservation* or through their *emulation*, as a practical solution for preservation of electronic documents.[19] The first approach could find a justification only in the context of cultural heritage preservation, where the *intrinsic* value of the document may justify the preservation of original decoding equipment;[20] the second option appears difficult to realize over large scale, both from an economic and from an software engineering perspective, and thus, seems doomed to remain confined to niche applications.[21]

V.2. Resignature

The EESSI consortium has also sought to address the need for ensuring the long-term integrity

[17] For example, in France, the *Fédération Nationale des Tiers de Confiancce* (www.fntc.org).

[18] [13] p. 34.

[19] For a complete review of available strategies for archival preservation of electronic documents, see [15].

[20] For example, the United States Constitution constitutes such a document with intrisic value goes beyond mere informational content.

[21] For example, emulation of videogames. Holland's Koninklijke Bibliotheek (www.kb.nl) has explored the practicality of emulation as a preservation strategy: see [16].

of cryptographically signed documents through its standard on "Electronic Signature Formats" [17]. The format distinguishes between two signature verification moments, "initial validation" and "late validation" (corresponding respectively to steps 2 and 4 of the signature lifecycle defined in section IV). The format for late validation encapsulates all of the information that can be eventually used in the validation process, such as revocation information, timestamps, signature policies, etc, while initial validation is used to gather this information in order to construct the late validation format.

However, the distinction between initial and late validation is founded on an analysis exclusively concerned with the security threat to signatures induced by decay in cryptographic strength: "Before the algorithms, keys and other cryptographic data used at the time the [electronic signature] was built become weak and the cryptographic functions become vulnerable, […] the signed data […] should be timestamped. If possible this should use stronger algorithms (or longer key lengths) than in the original timestamp. The timestamping process may be repeated every time the protection used to timestamp a previous [electronic signature] become weak."[22]

That is, the primary security concern here is modeled as one where advances in cryptanalysis could make it possible, some years after the moment of signature creation, to deduce the original private signing key. Cryptographic signatures would then no longer provide credible evidence suitable for litigation purposes, since such a scenario reproduces the conditions of a symmetric key cryptosystem — where signer and verifier both have access to the same key. To guard against this threat of decay, EESSI signatures are designed to be regularly timestamped afresh, with signing algorithms and key sizes appropriate to state-of-the-art cryptanalytic methods.

Such a solution does not address the problem of simultaneous preservation of legible documents and verifiable signatures. In fact, it further compounds it, encasing the bitstring underlying the electronic document in ever deepening layers of cryptographic signatures.

V.3. Canonicalization

Clifford Lynch proposed in 1999 the use of "canonical formats" as a preservation strategy for digital information [18]. Drawing on this approach, the Internet Engineering Task Force (IETF) has developed specifications to dealing with the issue of long-term preservation of cryptographic signatures founded on the use of canonical formats. In computer science, canonical refers to the process of conforming to an authoritative or authorized definition. In this case, canonicalization refers to the process of translating an encoded text into a version conformant with some canonical definition of that encoding.

The perceived usefulness of canonicalization for digital signatures is made clear in the case of the S/MIME secure messaging format, which defines the various data structures making it possible to cryptographically sign "plain text" email messages [19]. Unfortunately, there are

[22] [17], p. 16.

no universally adopted standards for representing plain text (ASCII and Unicode are standards for *character*, not *text* encoding) and because Windows, Mac, Unix platforms use different characters for indicating the end of a line,[23] a "plain text" email message will undergo a subtle and largely invisible transformation as it moves across computing platforms, a transformation that ensures, among other things, that lines are correctly terminated.

Such a transformation poses a very real problem for cryptographic signatures, which cannot tolerate *any* modification of the original message — even one involving a change of invisible characters. Thus, the S/MIME standard specifies that: "[e]ach MIME entity MUST be converted to a canonical form that is uniquely and unambiguously representable in the environment where the signature is created and the environment where the signature will be verified. […] The most common and important canonicalization is for text, which is often represented differently in different environments. MIME entities of major type "text" must have both their line endings and character set canonicalized."[24]

Thus, the S/MIME compliant sending agent processes the email message so that it conforms to the canonical encoding of plain text required by the standard. This will enable the receiving agent to adequately process the message and to verify the signature.

In practice, the effect of using canonical formats is to perform a format migration *before* the signature occurs, thus minimizing the effect of logical format decay. In this way, documents that have undergone canonicalization are less susceptible to simple transformations of the logical format (such as whitespace normalization), which immediately invalidate digital signatures. While this approach does address the problem of encoding format decay, it does nothing to eliminate it.

Thus, all three approaches share a fundamental assumption: *the authenticity of an electronic document is best ensured through the preservation of the integrity of the underlying bitstring*. Such a conception must be confronted with the strikingly different one adopted by the profession which, historically, has been entrusted with the social mission of preserving the integrity and intelligibility of documentary evidence, that of archivists.

VI. Archival Responses

Faced with either legislation granting special evidential value to digitally signed documents (European Union) or with government-wide PKI development projects (United States, Canada and Australia), archival institutions have had to determine how they would deal with cryptographically signed records. Several of them — among others, the National Archives and Records Administration (NARA), Library and Archives Canada, and the National Archives of Australia— have thus issued guidelines which seek to advise governmental agencies in the steps necessary to preserve records which may be digitally signed and may eventually be transferred into custody of archivists.

[23] Windows uses a two-character (carriage return + line feed) sequence, Macintosh uses a single carriage return, and UNIX uses a single line feed.

[24] [19], sec. 3.1.1

VI.1. US National Archives and Records Administration

If American federal or state legislation does not, overall, explicitly grant cryptographic signatures special status as evidence, the National Institute of Standards and Technology (NIST) is leading the development of a Federal Government Public Key Infrastructure, in coordination with industry and technical groups.[25] The National Archives and Records Administration thus issued in 2000 guidelines intended to help agencies expecting to produce, retain, and eventually transfer to NARA, digitally signed documents [20].

The guidelines suggest two distinct approaches to solving the problem of digital signature preservation: on the one hand, the agency may retain sufficient contextual information to adequately document the processes in place at the time the record was electronically signed. That is, "the agency's preserves the signature's validity and meets the adequacy of documentation requirements by retaining the contextual information that documented the validity of the electronic signature at the time the record was signed."[26] Such an approach is deemed more appropriate for records with long-term retention requirements, as it is less subject to the effects of technological obsolescence.

On the other hand, agencies may preserve the ability to validate signatures, that is, preserving both the contextual and structural information of the record, an approach NARA deems "potentially more burdensome, particularly for digitally-signed records with long retention needs, due to issues of hardware and software obsolescence."[27] The guidelines distinguish between the content, context and structure of electronic records, noting that "for an [electronic] record to remain reliable, authentic, […] it is necessary to preserve its content, context, and sometimes structure." Arguing that digital signatures are simultaneously part of the content, of the context, and of the structure of a digitally signed document, the guidelines conclude that in order to preserve the capacity to validate signatures, "it is necessary to maintain the structure of the electronic signature. In that is case, it is necessary to retain the hardware and software that created the signature (e.g., chips or encryption algorithms) so that the complete record could be validated at a later time."[28]

Whichever of the two approaches is chosen, NARA requires that for records to be permanently retained, "agencies must ensure that the printed name of the electronic signer, as well as the date when the signature was executed, be included as part of any human readable form (such as electronic display or printout) of the electronic record. NARA requires this so that the name of the signer will be preserved as part of the record."[29]

VI.2. National Archives of Australia

Since 2001, all administrative agencies of the Australian government must conform to

[25] See http://csrc.nist.gov/pki/.

[26] [20] p. 26.

[27] [20] p. 8.

[28] [20], p 10.

[29] [20], p. 33.

Gatekeeper® — a regulatory scheme framing the federal government PKI — in all cases where an electronic authentication system is required for the provision of governmental service. As a consequence, the National Archives of Australia published in May 2004 guidelines relative to the preservation of digitally signed documents [21]. The distinctive feature of the guidelines is to suggest that governmental agencies choose their preservation strategies based on a *risk analysis* of the likelihood that the document will be used in the context of litigation, and thus, the likelihood that the digital signature will need to be verified in the future.

If the risk of such an event is low or average, the guidelines suggest that agencies use metadata in order to record the existence and validity of the digital signature, including (a) the unique identifier of the relevant public key certificate, and that of the organization which produced it; (b) information relative to the digital signature associated with the document, e.g., the algorithm used to produce the signature; (c) information relative to the time and date when the digital signature was applied and/or verified with success.

If the risk of litigation is high, the guidelines recommend that administrative agencies implement a key management plan providing access to the full set of data necessary for signature verification for the full duration of document lifecycle. Such a plan must encompass the preservation of the public key certificates, of revocation lists, timestamps, and information relative to system audits.

In the specific case of documents that may eventually be transferred to the National Archives, the guidelines underline that "it is unlikely that there will be a continuing business need for any attached digital signatures to remain functional." Thus, "[t]he Archives will be unable to re-validate digital signatures attached to records because it will not attempt to gain possession of the relevant public and private keys (or equivalent device). … Why? It is impossible for the National Archives to gain access to and store all the components of authentication schemes necessary to ensure their ongoing functionality."[30]

VI.3. Library and Archives Canada

The 1999 Canadian Government Throne Speech announced an ambitious plan to make all federal programs and services available on-line by 2005. A key element of such a plan was to be the establishment of the "Government of Canada Public Key Infrastructure" project to meet the specific security requirements of federal electronic services delivery.[31]

Library and Archives Canada have thus issued guidelines relatives to the preservation of digitally signed documents [22], guidelines offering perhaps the bluntest assessment of the archival position with respect to the role of digital signatures in ensuring the evidential value of records: "For National Archives' purposes, the integrity and authenticity of records will continue to be inferred from their placement within an organization's record-keeping system

[30] [21], p. 36.

[31] See http:// http://www.solutions.gc.ca/pki-icp/. Equally importantly, the project was to provide a key market for the nascent Canadian PKI industry, in particular, Entrust, an offshoot of the now defunct Bell-Northern Research (see www.entrust.ca).

during the normal course of business, and from proof of that organization's reliance on records kept within their record-keeping system."

Such an assessment implies that, from the archivist's point of view, whatever security role digital signatures may have played prior to their transfer to the archives, they will by then have outlived their usefulness. Accordingly, "the National Archives will not attempt to maintain the capacity to re-verify a digital signature after transfer to its control, nor to preserve the traces of a digital signature generated under the current federal PKI system."

Thus, from the point of view of archival institutions confronted with the need to develop policies relative to the preservation of digitally signed documents, three possible solutions have emerged:

(1) *Preserve the digital signatures:* This solution supposes the deployment of considerable means to preserve the necessary mechanisms for validating the signatures, and does not address the need to simultaneously preserve the intelligibility of documents;

(2) *Eliminate the signatures:* This option requires the least adaptation from archival institution, but impoverishes the description of the document, as it eliminates the signature as one technical element used to ensure the authenticity of the documents;

(3) *Record the trace of the signatures as metadata:* This solution requires little technical means, and records both the existence of the signature and the result of its verification. However, digital signatures lose their special status as the primary form of evidence from which to infer the authenticity of the document.

While the first solution has often been implicitly codified in evidence law reforms (perhaps without realizing its full practical implications), it is the last solution which is most congruent with both archival practice and theory: "the findings of InterPARES indicate that integrity assurance and continuing accessibility are the key outputs of the archival recordkeeping function and that these are primarily assured through procedural and descriptive metadata. … Archival metadata must support the continued authenticity of records by describing the records as they were received from the records' creators and thoroughly documenting the entire process of preservation" [33].

VII. Conclusion

The gap between the responses offered by the legal, technical and archival community over the long-term preservation of digitally signed documents is best understood as a clash between two differing conceptions of electronic authenticity.

The first, espoused by the technical community and adopted by some segments of the legal community, is based on the measurement of a *physical* property of the document — *bitwise integrity* — whereby "data has not been altered in an unauthorized manner [i.e., by insertion, re-ordering, inversion, substitution, or deletion of bits] since the time it was created, transmitted, or stored by an authorized source"[23]. The appeal of such a measure lies in the hope that authenticity may become susceptible to precise quantification, to be given a simple thumbs up or down.

From the point of view of the archival mission, such a physical measure of authenticity is highly useful at specific points in the document lifecycle — for example, when transmitting documents across space. However, as the *primary* method for establishing authenticity, it effectively compounds the preservation problem.[32] Archivists prefer to rely on a second conception of electronic authenticity, one best described as *contextual*, which documents the totality of the controls and procedures, whether human or computer-based, that insure the identity and integrity of an electronic record throughout the totality of its lifecycle.[33]

The initial enthusiasm generated by cryptographic signatures, which led many to praise it as intrinsically superior to handwritten signatures,[34] is usefully compared alongside that generated by DNA profiling in criminal law. While this technology was initially granted a status of irrefutable proof of identification, it met with a surprising defeat during the course of the O.J. Simpson trial in 1995. As three sociologists of science explain, "[…] by following the samples from the crime scene to the laboratory, and then from the laboratory to the tribunal, one realizes that the genetic fingerprint may only serve its role of competent witness *if and only if* the succession of transactions during sampling, transport, preservation, digitization, and analysis of the sample is itself testified to by witnesses, certified and duly registered by responsible authorities. To be considered as such, the truth contained in the automatic signature (the genetic bar code) must be accompanied, surrounded by a whole series of bureaucratic traces: handwritten signatures on standard forms, actual bar-codes affixed on bags containing the samples, etc." [27].

It is those traces that were successfully contested during the Simpson trial, because, as archivists have long known, *no evidence is ever self-intelligible*. The same principle applies to electronic records: in order to be a "competent witness" of a juridical fact (commitment to obligations), an electronic document must be accompanied by traces of all of the operations which it is susceptible to incur: creation, modifications, annotations, signature, conversion, transmission, etc. Likewise, digital signatures are unable to testify *in and of themselves* of the identity and integrity of a document, and to be effective, must also be accompanied by the various traces that testify to their own identity and integrity as evidence — public key certificates, revocation lists, certificate chains, audit trails, hash fingerprints, etc.

The lesson here is that criteria for electronic authenticity will not be established by a technological silver bullet [28]. Just as signatures themselves were once technological novelties

[32] This is what the InterPARES research project expressed when declaring that "it is impossible to preserve an electronic record as stored physical object; it is only possible to preserve the means to make this document manifest" [24].

[33] Criteria for this type of context-based authenticity have been offered by the InterPARES research project as *benchmark* et *baseline requirements*. See [25].

[34] The best example of this line of thinking is offered in [8]: "Throughout history, lawmakers of both civil and commonlaw jurisdictions have sought rules that achieve the type and level of non-repudiation made possible by digital technology. Signatures, seals, notaries, recording offices, and certified mail are all examples of traditional mechanisms employed in efforts to supply and bolster non-repudiation. … Explicit consciousness of this powerful issue has surfaced only very recently, as society has faced the challenge of first matching and then exceeding traditional legal protections in the emerging digital communications environment." (564)

around which social practices gradually coalesced [29], the evidential value of electronic documents will emerge out of the slow and gradual engagement of relevant social groups with the various technical means supporting claims of authenticity. While legislation can provide a rich framework to support this engagement, efforts to dictate its precise rules are still premature at best.[35]

[35] For a more in-depth discussion on the idea of the socio-cultural foundations of evidence, see [30] and the more recent discussions by Xavier Lagarde [31, 32].

References

[1] Piette-Coudol, (T.), *La Signature électronique*, Paris, Littec, 2001.

[2] Renard, (I.), *Vive la signature électronique,* Paris, Delmas, 2002.

[3] Morin, (H.), Pourquoi la signature électronique reste lettre morte. *Le Monde*, 22 juin 2003.

[3] Kahn, (D.), *The Codebreakers,* New York, The Macmillan Company, 1967.

[4] Kahn, (D.), Cryptology Goes Public. *Foreign Affairs,* **58**, pp. 141-159, 1979.

[5] Landau, (S.), Primes, Codes and the National Security Agency. *Notices of the American Mathematical Society*, **30**, pp. 7-10, 1983.

[6] Landau, (S.), Zero Knowledge and the Department of Defense. *Notices of the American Mathematical Society*, **35**, pp. 5-12, 1988.

[7] Diffie, (W.), Hellman, (M. E.), New Directions in Cryptography. *IEEE Trans. on Inf. Th.,* **22** pp. 644–654, 1976.

[8] Ford, (W.), Baum, (M), *Secure Electronic Commerce: Building the Infrastructure for Digital Signatures and Encryption*, Upper Saddle River, NJ, Prentice Hall, 2000.

[9] EPIC, *Cryptography & Liberty 2000: An International Survey of Encryption Policy*, Washington: Electronic Privacy Information Center, 2000.

[10] ABA, *Digital Signature Guidelines*, Washington, American Bar Association, 1996.

[11] Meehan, (S. C.), Beard, (D.B.), What Hath Congress Wrought: E-Sign, The UETA, and the Question of Preemption. *Idaho L. Rev.*, **37**, pp. 389-414, 2001.

[12] Lenstra, (A.K.), Verheul, (E.R.), Selecting Cryptographic Key Sizes. *Journal of Cryptology,* **14**, pp. 255-293, 2001.

[13] Libon, (O.), Van Den Eynde (S.), *Trusted Archival Services*, European Electronic Signature Standardization Initiative, European Commission, 2000.

[14] Dumortier, (J.), Van Den Eyde, (S.), Electronic signatures and trusted archival services, *in Proceedings of the DLMForum 2002, Barcelona 6-8 May 2002*, Luxembourg, Office for Official Publications of the European Communities, pp. 520-524, 2002.

[15] Thibodeau, (K.), Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, *in The State of Digital Preservation: An International Perspective*, Washington D.C.: Council on Library and Information Ressources, 2002.

[16] Rothenberg, (J.), Ensuring the Longevity of Digital Documents, *Scientific American,* **272**, pp. 24–29, 1995.

[17] Pinkas, (D.), *Electronic Signature Formats*, European Electronic Signature Standardization Initiative, ETSI TS 101 733 V1.2.2.

[18] Lynch, (C.), Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information. *D-Lib Magazine,* **5**(9), 1999.

[19] IETF, *S/MIME Version 3 Message Specification – RFC 2633*, Internet Engineering Task Force,

1999.

[20] NARA, *Records Management Guidance for Agencies Implementing Electronic Signature Technologies*, Washington, National Archives and Records Administration 2000, available at: http://www.archives.gov/records-mgmt/policy/electronic-signature-technology.html.

[21] NAS, *Recordkeeping and Online Security Process : Guidelines for Managing Commonwealth Records Created or Received Using Authentication or Encryption*, Canberra, National Archives of Australia, 2004. Available at: http://www.naa.gov.au/recordkeeping/er/security.html.

[22] LAC, *Guidelines For Records Created Under a Public Key Infrastructure Using Encryption And Digital Signatures*, Ottawa, Library and Archives Canada, 2001. Available at http://www.collectionscanada.ca/06/0618_e.html.

[23] Menezes, (A.J.), van Oorschot, (P.C.), & Vanstone (S.A.), *Handbook of Applied Cryptography*, Boca Raton, FL, CRC Press, 1996.

[24] Duranti, (L.), *et al.*, Strategy Task Force Report, *in The Long-term Preservation of Authentic Electronic Records,* Vancouver, InterPARES, 2002

[26] MacNeil, (H.), *et al.*, Authenticity Task Force Report, *in The Long-term Preservation of Authentic Electronic Records,* Vancouver, InterPARES 2002.

[27] Lynch, (M.), McNally (R.), Daly, (P.), Le tribunal : Fragile espace de la preuve. *La Recherche*, **300**, pp. 112-115, 1997.

[28] Anderson, (R.), Why Cryptosystems Fail. *CACM*, **37**(11), pp. 32-40, 1994.

[29] Fraenkel, (B.), La *Signature: Genèse d'un signe,* Paris, Gallimard, 1992.

[30] Lévy-Bruhl, (H.), *La preuve judiciaire – Etude de sociologie juridique*, Paris, Librarie Marcel-Rivière et Cie, 1964.

[31] Lagarde, (X.), *Réflexion critique sur le droit de la preuve*, Paris, Librairie générale de droit et de jurisprudence, 1994.

[32] Lagarde, (X.), Vérité et légitimité dans le droit de la preuve. *Droits* **23**, pp. 31-39, 1996.

[33] Gilliland-Swetland (A.), Electronic Records Management. *ARIST* **39**, pp. 219-25, 2005.