

Les technologies de l'écrit électronique : synthèse et évaluation critique *

Jean-François BLANCHETTE †

Centre d'études sur la coopération juridique internationale
Centre national de la recherche scientifique

18 janvier 2001

* Ce document a été rédigé dans le cadre des travaux du groupe de réflexion sur les actes authentiques électroniques formé par le GIP Droit et Justice, et dont le rapport final a été remis au Ministère de la justice en juillet 2001. La rédaction de ce document a profité des apports de nombreuses personnes : Mme Isabelle de Lamberterie, CECOJI, Centre national de la recherche scientifique; Nick Mettyear, société PenOp; Johan Rommelaere, société LCI SmartPen; Graham Shaw, société Signum Technologies; Isabelle Guyon-Renard, François Frankel et Jacques Bluteau, Service central de l'état civil, Ministère des affaires étrangères; Me Menut et Me Bobant, Chambre nationale des huissiers de justice; M. Jean Berbineau, COMI, Ministère de la justice; Me Motel, Me Mathias, Me Delpeuch, Me Gard, M. de Martel, Mme Moreau-Bosc, Conseil supérieur du notariat; Me Lambert, M. Bouchon, M. Lemogne, Association pour le développement du service notarial; M. Gentilini, M. Henry-Bonniot, Tribunal de grande instance de Reims; Daniel Poulin, Pierre Trudel, Centre de recherche en droit public, Université de Montréal; Béatrice Fraenkel, Centre d'étude de l'écriture, Centre national de la recherche scientifique; Françoise Banat-Berger, Service des archives, Ministère de la justice. Ce rapport a été rendu possible par une bourse doctorale du Conseil de la recherche en sciences humaines du Canada et par une bourse de coopération France-Québec du Ministère des affaires internationales du Gouvernement du Québec. Toute erreur et/ou opinion n'engage évidemment que la seule responsabilité de l'auteur. © Jean-François Blanchette 2001.

† Centre de recherche en droit public, Faculté de droit, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7; Tél: 514-343-7533; Fax: 514-343-7508; Courriel: blanc@lexum.umontreal.ca.

A un homme muni d'un marteau, tout ressemble à un clou.
— Proverbe américain.

Table des matières

1	Introduction	5
2	Encodage	7
2.1	Grille d'évaluation des formats d'encodage	8
2.2	Le format texte	9
2.2.1	Exemple d'un fichier texte	9
2.3	Les formats Word et RTF	10
2.3.1	Exemple d'un fichier RTF	10
2.4	Le format HTML	11
2.4.1	Exemple d'un fichier HTML	11
2.5	Le format XML	12
2.6	Format de fichiers d'images	15
2.7	Les formats PostScript et PDF	16
2.7.1	Exemple d'un fichier PostScript	18
2.8	Pourquoi une norme?	18
2.9	Quel format d'encodage pour l'acte authentique électronique?	21
3	Signature	23
3.1	La signature cryptographique	24
3.1.1	La certification	26
3.1.2	Les infrastructures à clés publiques	27
3.1.3	Brèves remarques sur l'histoire de la cryptographie	28
3.1.4	Normalisation	29
3.1.5	Mesure de la sécurité	30
3.1.6	Évaluation	32
3.2	La signature biométrique	32
3.2.1	Mécanisme de la signature biométrique	34
3.2.2	Quantification, normalisation	35
3.2.3	Évaluation	36
3.3	la signature-tatouage	36
3.3.1	Mécanisme de la signature-tatouage	39
3.3.2	Quantification, normalisation	39
3.3.3	Sécurité	40
3.3.4	Évaluation	40
3.4	La signature numérisée	41

3.4.1	Évaluation	41
3.5	Conclusion	43
4	Archivage	47
4.1	Pourquoi la resignature?	49
4.2	L'approche EESSI	50
4.3	Conclusions	52
5	Conclusions	54

Chapitre 1

Introduction

À la mi-temps des travaux du groupe de travail sur l'acte authentique électronique, il apparaît utile de faire le point sur cet objet mystérieux. En effet, malgré les nombreuses et riches discussions, certains aspects de la problématique ont été peu abordés ; en particulier, on peut se demander à quoi ressemble *concrètement* cet acte authentique électronique : de quelle matière informatique est-il fait précisément ? Quelles sont les possibilités offertes et les contraintes imposées par les différentes technologies qui réalisent, dans le concret, l'acte authentique électronique ? Si on a remarqué à plusieurs reprises que le groupe de travail sur la dématérialisation des actes authentiques a du travailler à partir d'un texte de loi initialement conçu pour le commerce électronique, on a moins observé que nous avons également hérité des *technologies* issues du commerce électronique.

Le but de ce document est de donc présenter un survol des problématiques techniques particulières que pose l'acte authentique électronique, et d'offrir un survol aussi étendu que possible des technologies actuellement disponibles qui tentent d'y répondre. Un tel document permettra, il est espéré, de fonder les discussions juridiques sur une assise plus plus tangible, moins exclusivement théorique,¹ et de confronter directement les technologies impliquées dans la construction de l'acte authentique électronique : qu'accomplissent exactement ces technologies ? Quelles sont leur modalités de fonctionnement, d'utilisation, leur a-priori analytiques ? Que peut-on dire de leur déploiement à grande échelle et sur le long terme ?

Cette étude préliminaire ne prétend pas aucunement à l'exhaustif : l'ampleur du champ, la somme des informations techniques à synthétiser, et les moyens à ma disposition ne permettent tout simplement pas de produire un rapport qui tiendrait compte de l'ensemble des facteurs en jeu. L'objectif beaucoup plus modeste de ce rapport est plutôt d'ouvrir le champ de la discussion avant que celle-ci ne s'enferme dans des solutions technologiques trop précises. La question de la sécurisation des échanges électroniques est, à mon avis, à peine entamée, tant du point de vue technique que réglementaire, sans même parler de l'adoption de ces nouvelles façons de faire par les utilisateurs. Pourtant, selon certains, les solutions technologiques permettant d'assurer la sécurisation pérenne des documents existent déjà, ne resterait qu'à en assurer le déploiement... Comme nous pourrions le constater, la situation est, en réalité, beaucoup plus complexe et beaucoup moins

1. Un tel exercice a été brillamment réussi par M. Dominique PONSOT, dans son rapport *Valeur juridique des documents conservés sur support photographique ou numérique*, Observatoire juridique des technologies de l'information, 1995 — un rapport qui démontre la possibilité d'un langage qui soit tout à la fois intelligible au juriste et scientifiquement précis.

reluisante que les brochures commerciales satinées ne le laissent croire...

Le premier travail conceptuel important est de distinguer les différents étapes de la vie de l'acte authentique, et d'examiner les questions technologiques qui se posent à chacune de ces étapes. Nous distinguerons les étapes suivantes dans la vie de l'acte, avec les questions afférentes :

- 1° **sa rédaction** : logiciel de rédaction ; format d'encodage de l'acte ;
- 2° **sa signature** : type de technologie ; outil de signature ; présence des parties à la signature ; signification de la signature de chacun des parties ; formalisme de la signature ; mécanismes d'intégrité additionnels ;
- 3° **la transmission du document** : stockage initial ; mécanisme de transmission ; nature de l'émetteur et du destinataire ;
- 4° **la réception du document** : notification du destinataire ; délais de transmission ; intelligibilité du document ; vérification des signatures ;
- 5° **le stockage à long terme** : format d'encodage ; support ; mécanismes institutionnels ; production des copies conformes ; vérification de la signature ; valeur probante.

Il est difficile de traiter de ces rubriques indépendamment les unes des autres et, idéalement, ce rapport les traiterai toutes, mais, faute de moyens et de temps, nous découperons la vie de l'acte en trois parties : l'encodage, (chapitre deux), la signature (chapitre trois), et l'archivage (chapitre quatre), pour conclure (chapitre cinq) par quelques observations générales sur l'acte authentique électronique et sa sécurisation.

Chapitre 2

Encodage

Au sein d'un système informatique, toute information est nécessairement représentée sous forme d'un encodage. Les publicités expriment souvent cette réalité en saupoudrant les images de 0 et de 1, mais en fait, l'encodage binaire n'est utilisé que dans les couches matérielles basses de l'ordinateur. Les applications informatiques définissent plutôt des encodages de haut niveau, encodages permettant d'entreposer de l'information dans un fichier de façon à pouvoir, d'une part, la conserver, d'autre part, la transmettre et l'échanger. Ainsi, tout traitement de texte offre une fonction de sauvegarde d'un document de travail sous forme d'un fichier, de façon à pouvoir travailler sur un même document en plusieurs étapes, et d'autre part, de façon à pouvoir échanger le fichier avec d'autres utilisateurs.¹ L'encodage est la manière dont cette information est structurée au sein du fichier — nous verrons plus loin des exemples qui permettront de rendre cette notion plus concrète.

Si le besoin initial pour un format d'encodage — entreposer, transmettre, échanger — est relativement évident, il est beaucoup plus subtil de comprendre pourquoi différents types d'encodages sont apparus, se sont imposés ou, le plus souvent, disparus sans laisser de trace. Pour ce faire, il faut explorer les logiques techniques, économiques et historiques sous-tendant ces formats.

Nous nous intéressons ici uniquement à l'encodage de documents. Évidemment, l'objet « document » peut être compris de différentes façons : contenu sémantique, apparence visuelle, série de caractères alphanumériques, collection d'informations, etc. Les normes d'encodage des documents privilégient en général un seul de ces aspects, et intègrent les autres de façon secondaire. Nous verrons que les questions relatives à l'encodage se posent à chacune des étapes de la manipulation de l'acte authentique électronique : à l'établissement évidemment, mais aussi à la signature (problème du *what you see is what you sign*) et à l'archivage (problème de la migration des fichiers).

Il existe à peu près autant d'encodages que d'applications.² Certains sont totalement inconnus du grand public, alors même qu'ils dominent au sein de communautés d'utilisateurs.³ L'encodage

1. La gestion — l'accès, destruction, etc. — des fichiers est réalisée au niveau du système d'exploitation, mais l'encodage du fichier est fonction de l'application informatique.

2. La bible des formats d'encodage graphique : James D. Murray et William van Rypper, *Encyclopedia of Graphics File Formats, 2nd Edition*, O'Reilly, 1996.

3. Par exemple, les mathématiciens utilisent le langage T_EX pour produire des documents en format DVI, offrant un typographie de qualité supérieure, tout en étant relativement indépendant de la plate-forme informatique. La difficulté

peut varier selon le type d'application l'utilisant (traitement de texte, traitement d'image, base de données), la communauté d'utilisateurs visée, etc. L'encodage peut faire l'objet d'une norme ou non — nous traiterons de la question de la normalisation des encodages à la fin de cette section.

En considérant la notion juridique de « lisibilité » dans le contexte des documents électroniques, il faut garder à l'esprit que tout encodage est inclus dans un chaîne d'autres encodages. Le simple affichage d'une lettre à l'écran est le fruit d'une interaction touffue de ces encodages : encodage des caractères (ASCII, Unicode), polices de caractères (PostScript, TrueType), structuration de l'information sous forme d'un document (Word, Wordperfect), structuration de l'information (XML, etc), couleur (ColorSync, etc), modèle d'imagerie (GDI, Quickdraw, etc). Chacun de ces encodages évolue selon une dynamique qui lui est propre et qui varie selon des facteurs économiques, problématiques techniques, influences historiques. De plus, un format d'encodage ne peut être compris que comme un élément donné dans la chaîne totale des équipements qui le rend intelligible : un fichier Word ou HTML est toujours conçu pour être jumelé au logiciel Word ou à un fureteur Web, lui-même conçu pour un certain modèle d'ordinateur et un certain système d'exploitation. Du point de vue du groupe de travail, ceci souligne que la notion de *lisibilité d'un document électronique ne peut être comprise en dehors de l'interaction de l'ensemble de ces encodages avec le logiciel et le matériel informatique conçu pour les interpréter.*

2.1 Grille d'évaluation des formats d'encodage

Dans le contexte de la sélection d'un encodage approprié à l'acte authentique électronique, il faut considérer un ensemble de questions :

- 1° **Composition** : Peut-on composer le document à partir de bases de données existantes ? Comment peut-on créer le document à partir d'archives papier ? Peut-on le créer à partir d'un document à la volée ?
- 2° **Extraction** : Comment peut-on extraire de l'information structurée du document ?
- 3° **Forme et fond** : L'encodage donne-t-il une forme fixe au document (ou, préservation des qualités visuelles du document ?) L'encodage respecte-t-il la forme du document ? La présentation du document varie-t-elle selon les équipements ?
- 4° **Signature** : Quel type de signature l'encodage permet-il ? (cryptographique, biométrique, stéganographique, image ?)
- 5° **Modifications** : L'encodage permet-il de faire des ajouts, modifications, annotations ? Comment ces annotations sont-elles archivées ?
- 6° **Normalisation** : Cet encodage est-il normalisé ? Quelle est la nature de cette norme ? Quelles sont les institutions qui ont créées cette norme ? Quelles sont les perspectives de pérennité de cette norme ?
- 7° **Lisibilité** : Comment l'encodage est-il couplé à un outil de lecture ?
- 8° **Traductibilité** : L'encodage est-il susceptible d'être traduit en un autre encodage ?

d'apprentissage de l'encodage T_EX en fait un candidat peu probable comme format universel, mais pour les mathématiciens, c'est un outil indispensable et très désirable, qui a connu un succès fulgurant en moins d'une dizaine d'années. Voir à ce sujet Donald KNUTH, *Digital Typography*, CSLI Publications, 2000.

Encore une fois, le manque de temps et de moyens ne me permettront pas de traiter systématiquement de chacun de ces points individuellement. Dans le cadre qui nous concerne, on peut considérer les encodages suivants : 1°) texte « brut » ; 2°) Word et RTF ; 3°) HTML ; 4°) XML ; 5°) PostScript et PDF ; 6°) JPEG et TIFF. Chacun de ces formats est représentatif d'une façon particulière de concevoir la notion de document, le plus souvent à l'exclusion des autres notions. Pour chacun des exemples, j'ai tenté de fournir une représentation de l'encodage, quand c'était possible (ce n'est pas le cas pour un fichier TIFF), de façon à rendre la notion d'encodage moins abstraite. Une telle « cuisine » informatique n'est pas sans utilité pour mieux comprendre les défis que représente concrètement l'authenticité électronique. ⁴

2.2 Le format texte

C'est le niveau zéro de l'encodage d'un document, puisque tous les éditeurs de texte permettent d'afficher, lire, et imprimer des documents en format « texte ». Ce qui ne signifie par pour autant que ce soit un format universel, puisque chacun des trois systèmes d'exploitation dominants — Windows, Macintosh et Unix — codifient le format texte différemment, en utilisant des marqueurs de fin de ligne différents. En format texte, peu d'informations sont gérées : le format se contente de gérer les caractères alphanumériques, ainsi qu'un certain nombre de caractères « blancs » — retour de chariot, fin de ligne, espace blanc, espace de tabulation, fin de fichier, etc. Le plus souvent, ces caractères sont codés en ASCII, mais ce codage ne spécifiant pas la représentation des caractères accentués, il sera graduellement remplacé par le standard UNICODE. ⁵

Au sein de la communauté informatique, le format texte jouit d'une grande popularité pour la création de documents. Le plus souvent, les programmes informatiques sont écrits à l'aide d'éditeurs de texte programmables extrêmement sophistiqués (Emacs, Alpha, etc). Si le format ne permet pas de jouir d'aides visuelles comme le changements de polices ou de style au sein d'un document, ces éditeurs étagent les lignes et colorent les mots-clés d'un programme, de façon à le rendre plus facilement intelligible. Dans ce cas particulièrement, le texte devient véritablement fonction de l'interaction dynamique entre logiciel de lecture et fichier.

2.2.1 Exemple d'un fichier texte

Longue vie à l'acte authentique électronique!

On voit qu'aucune information particulière n'est conservé quant à l'apparence du document — pas d'informations sur la police de caractère, la taille, l'italique ou le gras, etc. On gère les espaces blancs et les caractères, c'est tout. C'est cette simplicité qui lui permet cependant une grande universalité, mais comme pour tout les formats d'encodage, c'est l'interaction entre du matériel, du logiciel, et un document qui permet d'afficher et d'imprimer cette phrase.

4. Pas nécessaire d'être diplômé en chimie pour réussir une génoise mais encore faut-il en connaître un peu sur l'interaction entre les œufs, la farine, l'air et la chaleur.

5. Ce standard vise à pouvoir encoder l'ensemble des alphabets, permettant d'encoder plus de 65 000 caractères différents, incluant l'ensemble des idéogrammes chinois, japonais, et coréens — voir <http://www.unicode.org>.

2.3 Les formats Word et RTF

L'encodage des fichiers créés par les différentes versions des logiciels de traitement de texte Word est le meilleur exemple d'une norme qui s'est imposée *de facto*. Bien que cet encodage aie été entièrement conçu par Microsoft, il s'est imposé du seul fait que le logiciel lui-même s'est imposé. Ainsi, il est courant d'échanger des documents en format Word, car la plupart des utilisateurs assument que tous et chacun disposent du logiciel pour les lire. Il existe aussi des logiciels qui permettent uniquement de lire des fichiers Word, sans pouvoir les modifier.⁶

Le format Word est avant tout conçu pour répondre à des besoins d'éditions modestes et variés, allant de la rédaction d'un thèse à la celle de lettres. Le logiciel est conçu dans une optique où le document est tout simplement imprimé sur papier, et son paradigme dominant est ainsi le fameux WYSIWYG, *What You See Is What You Get*, c'est-à-dire que le logiciel donne au texte la même apparence à l'écran que celle qu'il aura sur papier.

Word permet également de structurer l'information de façon assez sophistiquée. Cela a commencé avec la fonction de gestion des envois postaux de masse, mais ces fonctions sont exploitées par les logiciels de rédactions d'actes notariés, par exemple, pour échanger de l'information entre le logiciel et l'acte en cours de rédaction. Cette structuration de l'information ne fait l'objet d'aucune norme, elle est simplement interne à Word.

Il existe un format créé pour faciliter l'échange entre les différents logiciels de traitement de texte (Word, WordPerfect, etc), le format RTF (*Rich Text Format*), comme son nom l'indique, est un format texte mais qui spécifie des paramètres quant à la police, etc. Ce format a été élaboré par Microsoft pour permettre l'échange de fichiers entre les différents traitements de texte. Ce format n'est jamais rédigé directement, mais plutôt produit par les logiciels de traitement de texte, et tous les logiciels commerciaux le lisent, si ce n'est pour assurer un minimum d'interopérabilité avec le joueur dominant du marché — Microsoft Word.

2.3.1 Exemple d'un fichier RTF

```
{\rtf1\mac\deff2
{\fonttbl{\f2\froman New York;}
{\f20\froman Times;}
{\f22\modern Courier;}
{\stylesheet{\f16 \sbasedon222\snext0Normal;}}
{\info}\paperw11900\paperh16840\margl1701\marginr1701\margt1417\margb1417
\deftab709\widowctrl\ftnbj\sectd \sbknone\linemod0\linex0\headery1077
\footery1077\cols1\colsx709\endnhere\pard\plain \qc \f16
{\f20\fs96 Longue vie \ '88 l\quote acte authentique \ '8electronique!}
{\f2010\fs96 \par }}
```

Les informations du fichier RTF permettent donc à chaque application de reconstruire le document, incluant les notes de renvois, de gérer les caractères accentués, etc. Cette traduction n'est pas toujours parfaite, dans le cas de documents très complexes.

6. Voir Icword pour le Macintosh (<http://www.icword.com>), Microsoft Reader pour Windows (<http://www.microsoft.com/reader/>), pour Windows, et Ted (format RTF) pour Unix (<http://www.nllgg.nl/Ted/>).

2.4 Le format HTML

Le HTML (*Hyper-Text Markup Language*) est évidemment le langage développé pour le Web — c'est celui que les fureteurs *Netscape Navigator*, *Microsoft Explorer* et autres peuvent lire et traduire de façon à afficher des pages Web. Ce langage est fortement inspiré du SGML, langage développé par l'armée américaine dans le but de gérer la documentation des équipements militaires — nous reparlerons des caractéristiques du SGML dans la section traitant du XML. La philosophie ayant guidé la conception du HTML était celle de définir un langage simple permettant de créer rapidement des documents « hypertextes », c'est-à-dire intégrant des liens sur des objets distribués sur l'ensemble du réseau — autres pages, images, documents, etc. Un document n'est donc plus une entité fixe locale, mais devient un assemblage dynamique d'éléments distribués au travers du réseau.

Sa grande simplicité a beaucoup contribué à sa popularité, simplicité qui a permis à nombres de personnes sans formation technique particulière de pouvoir publier sur la toile, sans avoir à absorber des masses d'informations techniques indigestes — on peut créer une page Web en quelques heures à peine, « à la main », ou à l'aide de logiciels conçus à cet effet.

La définition du langage est sous la responsabilité du *World-Wide Web Consortium* (W3C) et cette définition a été l'objet de guerres rangées entre les fabricants (*Netscape* et *Microsoft*) qui ont tenté de créer des marchés pour leur fureteur — un exemple criant où un encodage est devenu un enjeu stratégique crucial. Cette guerre semble actuellement perdre de son importance, alors que le HTML semble vouloir être ultimement être remplacé par le XML.⁷

Le HTML effectue une certaine structuration des données à l'aide de « marques ». Par exemple, dans l'exemple ci-dessous, on peut diviser le document en deux parties, d'une part, l'en-tête, incluse entre les marques `<HEAD>` et `</HEAD>`, et d'autre part, le corps du document, inclus entre les marques `<BODY>` et `</BODY>`. Ces régions du texte sont ainsi structurées et on peut leur appliquer des traitements distincts. Cependant, cette structuration est imparfaite, pour deux raisons : d'une part, les utilisateurs ne peuvent définir une structuration qui leur est propre, et doivent nécessairement utiliser celle définie par le langage, et d'autre part, cette structuration mêle la forme et le fond, c'est-à-dire que le fureteur associe à une structure donnée une représentation donnée. C'est à ce niveau que le XML tente de corriger les faiblesses du HTML.

2.4.1 Exemple d'un fichier HTML

```
<HTML>
  <HEAD>
    <META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
    <META NAME="Generator" CONTENT="Microsoft Word 98">
  </HEAD>
```

7. Il est cependant à parier que Microsoft développera ses propres modifications au standard XML, de façon à en faire un outil stratégique de structuration du marché. Ce type de logique est particulièrement tenace chez ce fabricant et les premiers signes sont déjà dans l'air : voir Margret JOHNSTON « XML Factions Develop Along Familiar Lines » *InfoWorld.com*, 14 décembre 2000, qui note que « si Microsoft est largement en avance sur ses compétiteurs dans le marché du XML, la société n'entretient pas une relation exceptionnellement ouverte avec les gens qui définissent les mécanismes de la conversation ».

```

<BODY>
  <FONT FACE="Times" SIZE=7>
    <P ALIGN="CENTER">
      <A HREF="http://www.internet.gouv.fr/pubs/acte-authentique.html">
        Longue vie &agrave; l&#146;acte authentique &eacute;lectronique!
      </A>
    </P>
  </FONT>
</BODY>
</HTML>

```

Le document fait ici référence à un autre document sur la Toile, à travers le mécanisme des liens hypertextes. Les marques `<A>` et `` créent un lien entre le document situé à l'adresse `http://www...acte-authentique.html` et le texte « Longue vie à l'acte authentique électronique! ». Ainsi, lorsque l'on clique sur ce texte, le fureteur localise le document indiqué dans le lien et le charge à l'écran.

On voit aussi ici la façon dont différents type de normes sont intégrés au sein du format HTML. Par exemple, la référence à `CONTENT="text/html; charset-iso-8859-1"` fait référence à la norme multimédia de types MIME, qui permet à différents logiciels (principalement, les logiciels de courriels et les fureteurs Web) de gérer correctement les différents types d'objets que ces applications rencontrent. Dans ce cas-ci, un fureteur sait qu'il a affaire à un fichier HTML, de type « texte », que ce texte est encodé en ISO-8859-1, ce qui permet de choisir l'affichage approprié pour les caractères accentués. On voit bien ici comment les différentes normes sont imbriqués les unes dans les autres. La référence à `à` permet d'encoder l'accent grave de façon à ce qu'il soit portable entre les différentes plate-formes. De même, l'apparence du texte est déterminé par `FONT FACE="TIMES" SIZE=7` fait référence à une autre norme — celle définissant la police de caractère Times, et une taille, 7.

2.5 Le format XML

Le XML est une norme relativement récente, mais qui connaît actuellement une progression fulgurante. Le XML veut remplacer le HTML comme langage du Web, celui-ci n'ayant jamais été conçu pour répondre aux besoins auquel il fait face aujourd'hui — le commerce électronique entre autres. Le XML est un sous-ensemble épuré du SGML, adapté aux exigences du Web.⁸ La finalité du langage en est une d'*échange de données*, mais plus encore, de remplacer complètement tous les formats propriétaires de traitements de texte et d'édition électronique. A long terme, les concepteurs du langage désirent que tous ces programmes produisent du XML et du XSL (*eXtensible Style Language*), dont la combinaison pourrait répondre à tous les besoins imaginables d'édition, quelque soit la plate-forme informatique, quelque soit l'application.⁹

8. Un tel processus de simplification n'est pas inhabituel en informatique — ainsi, un des langages les plus utilisés en intelligence artificielle, le LISP a incorporé au fil des années tant d'ajouts que sa spécification dépasse les 1000 pages. Le langage Scheme incorpore les principales caractéristiques du LISP, mais sa spécification ne dépasse pas les quarante pages.

9. Voir Bosak, "Four myths about XML", disponible sur le site `www.w3c.org/`.

Le XML est une norme promulguée et gérée par le W3C (*World-Wide Web Consortium*), organisme fondé par Tim Berners-Lee, un des inventeurs du Web.¹⁰ Le W3C est, avec l'IETF, une de ces nouvelles institutions qui exercent une influence considérable sur le développement des normes techniques du Web et de l'Internet. Est-ce dire que le XML est un standard « ouvert »? On peut dire qu'il n'est pas propriétaire, au sens où aucune société ne contrôle directement la définition de la norme. Cependant, ceci ne signifie pas que tout un chacun peut participer au processus de rédaction de la norme — les membres du W3C sont principalement des corporations, le membership se payant à coup de \$50 000 dollars.

Un document XML est en fait composé de trois parties : il faut premièrement définir les catégories qui seront utilisées pour coder les données, au sein d'un DTD, *Document Type Definition*. Alors que cette catégorisation est implicite en HTML et ne peut être étendue par l'utilisateur, Le DTD permet de définir au gré de l'utilisateur la structuration des données. Par exemple, supposons qu'un usage désire définir une catégorie de données « Nom de jeune fille », catégorie susceptible d'être utilisée pour un acte juridique. La catégorie serait définie au sein d'un DTD, selon les règles du langage XML :

```
<!DOCTYPE acte-authentique>[
<!ELEMENT acte-authentique (notaire, parties*, acte)>
<!ELEMENT notaire (nom, adresse)>
<!ELEMENT nom (famille, prenom, nom-de-jf)>
<!ELEMENT famille (#PCDATA)>
<!ELEMENT prenom (#PCDATA)>
<!ELEMENT nom-de-jf (#PCDATA)>
<!ELEMENT parties ... <!ELEMENT acte ...]>
```

Cette DDT fictive et simplifiée définit une structure *acte-authentique*, qui contient des éléments *notaire*, *parties*, et *acte*. L'élément *notaire* est lui même défini comme étant constitué de la composition d'un élément *famille* et *prenom*, ces éléments étant eux-mêmes définis comme correspondants à des chaînes de caractères (#PCDATA).

Le document XML lui-même contiendra les données de l'acte authentique, données qui ne seront interprétables qu'en présence du DTD. Un document XML peut aussi contenir un lien sur un DTD, à la façon d'un lien hypertexte. Ainsi, l'interprétation d'un document XML peut être fonction d'un document non-local, situé ailleurs sur le réseau.

```
<acte-authentique>
  <notaire>
    <nom>
      <famille>Gard</famille>
      <prenom>Martine</prenom>
      <nom-de-jf>Gagné</nom-de-jf>
    </nom>
  </notaire>
```

10. Voir le site www.w3c.org.

```

    <parties>
    ...
  </parties>
  <acte>
  ...
  </acte>
</acte-authentique>

```

L'apparence du document est déterminée par un troisième document, une feuille de style exprimée en XSL.¹¹ Cette feuille de style a pour but de complètement dissocier l'apparence du document des données contenues dans le document. Ainsi, on peut associer à une donnée structurée une certaine apparence. Dans l'exemple qui suit, la feuille de style traduit un document XML en un document HTML. Pour chaque catégorie de données définie dans le DTD, on définit un patron (« template ») qui traite la représentation de chaque catégorie. Le template `acte-authentique` définit un document HTML dont la section `<body>` comprendra les trois catégories `notaire`, `parties`, et `acte`. La catégorie `notaire` est traitée par un autre patron, qui s'occupe d'aligner le nom et le prénom du notaire, et de les précéder par « Maître ».

```

<xsl:template match="acte-authentique">
  <html>
    <head><title>Un acte authentique</title></head>
    <body>
      <xsl:apply-templates select="notaire"/>
      <xsl:apply-templates select="parties"/>
      <xsl:apply-templates select="acte"/>
    </body>
  </html>
</xsl:template>

<xsl:template match="notaire">
  Maître
  <xsl:value-of select="famille"/>,
  <xsl:value-of select="prénom"/><P>
  <xsl:value-of select="nom-de-jf"/>,
</xsl:template>

```

Les feuilles de style permettent d'adapter la présentation d'un document XML au périphérique de visualisation : pour un même document, une feuille de style pourra ajuster l'affichage à l'écran d'un portable, d'un ordinateur personnel, ou d'un kiosque Internet public. Ce découplage entre forme et fond représente un des atouts majeur du XML, mais évidemment, le document n'est pas — ne peut être — affiché identiquement dans chaque cas !

On voit ainsi que le format XML met en place une machinerie complexe, et que le résultat final, tant à l'écran qu'à l'impression, dépend de l'interaction de plusieurs fichiers, possiblement distribués à différents endroits sur le réseau, de même que de la mécanique d'interprétation contenue dans le logiciel de lecture.

11. A noter que la spécification du XSL n'est pas encore, à ce jour, complétée.

2.6 Format de fichiers d'images

Un format image (on utilise parfois les termes *raster* ou *bitmap*) est un format qui encode une image selon un quadrillage dont la densité définit la résolution de l'image. Pour mieux comprendre cette définition, il est utile de saisir comment fonctionne trois périphériques informatiques essentiels au document numérique : l'écran, l'imprimante, et le scanner.

L'écran d'un ordinateur est composée de centaines de milliers de points lumineux, les *pixels*. La densité des pixels à l'écran varie selon les plates-formes : le Macintosh, par exemple, assigne 72 pixels à chaque pouce d'écran et on dit qu'il a une résolution de 72 points par pouce (ppp). Le pixel est susceptible de variations de couleur (au minimum, le noir et le blanc), et d'intensité lumineuse. A chaque pixel est assigné un certain nombre de bits — l'unité d'information binaire. Pour un écran noir et blanc, un seul bit est nécessaire (0 = pixel éteint, 1 = pixel allumé), alors que pour un écran couleur, le nombre de couleurs détermine le nombre de bits nécessaire (256 couleurs = 8 bits, 16 millions = 24 bits).

Une imprimante laser fonctionne d'une façon similaire : elle divise aussi la surface de la feuille de papier en un quadrillage de minuscules points (d'encre) , mais ce quadrillage est plus dense que celui d'un écran d'ordinateur — une imprimante laser typique assigne 300 ou 600 points par pouce, alors qu'une imprimante de qualité industrielle (RIP) peut atteindre une résolution de plus de 3300 ppp.

De même pour un scanner : il échantillonne la surface (analogue) d'un document en une image (numérique) d'une certaine résolution. Le scanner lui aussi définit un quadrillage dont la densité est définie par l'utilisateur et/ou par les capacités du scanner. Plus le quadrillage est dense, plus le fichier résultant de la numérisation sera gros : un fichier numérisé à 72 ppp sera 25 fois moins volumineux qu'un autre numérisé à 360 ppp, puisque chaque pouce carré de l'image sera échantillonné selon un quadrillage 25 fois moins dense. Ainsi, il y a un rapport direct entre la taille d'un fichier image et sa résolution.

La résolution finale d'une image est fonction à la fois de sa propre résolution et de celle du périphérique de visualisation : on peut très bien afficher une image de résolution 300 ppp sur un écran de 72 ppp, mais l'image ne pourra avoir plus de 72 ppp de résolution à l'écran. De même, on peut très bien imprimer une image de résolution 72 ppp sur une imprimante de 300 ppp, elle n'aura toujours que 72 ppp de résolution.

L'intérêt du format image, c'est qu'il reproduit, justement, une image. Les caractéristiques visuelles d'un document sont préservés : mise en page, marques manuscrites — la signature, par exemple. De plus, c'est la façon la plus simple de numériser les archives papiers, en obtenant tout simplement une image électronique de chaque page d'un document, et en liant ces pages par un mécanisme d'indexation approprié. Le format image permet ainsi de joindre l'univers papier et l'univers électronique. Cependant, un format image n'assigne aucune structuration sémantique aux données contenues dans l'image. En particulier, le texte écrit est compris en tant qu'image, et non en tant que suite de caractères intelligibles. Pour que l'ordinateur soit en mesure d'effectuer des traitements sur le texte contenu dans un fichier image, il faut effectuer une opération ultérieure

de reconnaissance optique de caractères.

Il existe des centaines de formats d'image : parmi les plus connus le format JPEG, format produit par ISO et l'ITU, et le format TIFF (*Tagged Image File Format*), dont la norme appartient à présent à la société Adobe.¹²

2.7 Les formats PostScript et PDF

Un des problèmes des fichiers image que l'on vient de décrire, c'est que l'on doit décider une fois pour toute de la résolution du fichier que l'on crée. Comme on l'a vu, rien ne sert d'imprimer un fichier de résolution 72 ppp sur une imprimante de 2400 ppp. Il faut donc conserver les fichiers à la meilleure résolution possible, mais ceci implique des fichiers de grande taille. Le langage PostScript propose une solution pour résoudre ce problème, en décrivant des images d'une façon uniforme quelle soit la résolution du périphérique d'impression, à l'aide d'équations vectorielles. Ainsi, dans un chaîne de production graphique, on peut disposer d'imprimantes de résolution très différentes connectées au même ordinateur — 300, 600 ppp, ou encore une Linotronic à même de produire du film en quatre couleurs, à des résolutions de 3386 ppp. Un document décrit en langage PostScript peut être envoyé à tout périphérique qui supporte le langage PostScript, quelque soit sa résolution, en produisant un résultat adapté à chaque type de périphérique. De plus, pour des documents de texte, le fichier PostScript correspondant peut être très compact, puisque d'une part, les polices de caractères PostScript les plus courantes sont conservées en mémoire directement sur l'imprimante, et d'autre part, qu'elles sont décrites sous formes d'équations vectorielles.

Le PostScript est un langage informatique de description de page (*page description language*) — c'est-à-dire que c'est un langage de programmation dont la seule finalité est de décrire des pages. Le langage est produit par un pilote d'impression, qui agit comme une interface entre l'ordinateur et l'imprimante. Le pilote produit un fichier PostScript qui est ensuite envoyé à un *interprète PostScript* résidant sur le périphérique d'impression. Cet interprète (qui prend la forme d'une puce dédiée) effectue la traduction entre les commandes contenues dans le fichier, et les caractéristiques matérielles du périphériques (résolution, taille du papier, couleur, etc.). Ce mécanisme permet donc d'effectuer la traduction entre d'une part, des plates-formes informatiques différentes (Macintosh, Windows, etc), et, d'autre part, des périphériques d'impression extrêmement variés.¹³

Le PostScript a été à l'origine d'une petite révolution informatique qui a fortement contribué à l'essor du micro-ordinateur, en permettant à tout un chacun de disposer d'outils compatibles de mise en page et d'impression de qualité professionnelle, et cette norme est aujourd'hui presque universelle au niveau des imprimantes. Il est aussi possible de représenter des fichiers PostScript à l'écran, à travers des applications comme Ghostview, qui traduisent un fichier PostScript dans le

12. Les services du Journal officiel, par exemple, numérisent chacune des pages du JO sous forme de document TIFF. Ainsi, le texte de la loi du 13 mars 2000 est disponible à l'URL <http://tif.journal-officiel.gouv.fr/2000/03968001.tif>.

13. A une certaine époque, l'installation d'un traitement de texte sur PC exigeait au fournisseur de fournir des pilotes d'impression pour *toutes* les imprimantes susceptibles d'être connectées à l'ordinateur (des centaines). Avec le PostScript, un seul pilote est nécessaire et il suffit de développer un interprète PostScript pour chaque périphérique d'impression.

modèle d'image utilisé pour Windows (GDI) et Macintosh (QuickDraw). La défunte plate-forme d'ordinateur NeXT avait adoptée une norme d'affichage Display PostScript, mais cette norme n'a jamais rencontré, de loin, le succès qu'a eu le PostScript au niveau de l'impression.

Le format PDF est basé sur le langage PostScript, et reprend sa philosophie. Alors que l'objectif du PostScript est de pouvoir disposer représenter une page de manière qui soit indépendante de la résolution du périphérique d'impression, l'objectif du PDF est de pouvoir représenter un document de la même manière, indépendamment de la plate-forme et de l'ordinateur utilisé — un document devrait apparaître de la même façon sur un Macintosh, un PC ou une station Unix. Ce problème est loin d'être trivial, puisque ces plates-formes différents dans nombres d'aspects qui influencent directement la représentation d'un document à l'écran : encodage des caractères accentués, standard des polices de caractères (PostScript ou True Type), etc.

Le format PDF peut-être produit à partir de tout logiciel, en utilisant un logiciel qui se substitue à un pilote d'impression : toute application susceptible d'imprimer peut ainsi produire du PDF, de la même façon qu'elle peut produire du PostScript. Le PDF se situe donc à un niveau plus intermédiaire dans la chaîne de production d'un document : on n'écrit pas un document directement en PDF, on produit le document dans le logiciel approprié et ensuite, on produit du PDF à partir de ce logiciel. C'est cette position intermédiaire qui donne au PDF sa grande inter-opérabilité.

A partir d'un document papier, plusieurs stratégies de capture sont possibles : on peut produire un document PDF qui ne contient qu'une image numérique du document, ou appliquer des outils de reconnaissance automatique de caractères, de façon à produire un document qui contient à la fois une image numérique du document et une représentation sémantique du texte, susceptible d'être indexée. Ce format permet ainsi de conserver une représentation visuelle exacte du document tout en offrant la possibilité d'effectuer des recherches dans le texte — un encodage qui tente de mitiger les inconvénients du format image en conservant une certaine « intelligence » au texte.

La société Adobe a habilement manœuvré de façon à promouvoir l'utilisation du PDF comme norme d'échange de documents sur la Toile : tout d'abord, le logiciel de lecture du format PDF est gratuit, facilement disponible, et il a été développé pour toutes les plate-formes — Windows, Macintosh, toutes saveurs d'UNIX, etc. Ensuite, les logiciels Acrobat Capture permet de numériser des documents papier en des documents PDF et en conservant au maximum l'apparence et la mise en page des documents. Le format incorpore en plus des aspects interactifs : liens hypertextes, signets, signature électronique, chiffrement, etc. Au-delà d'un simple encodage de documents, le format PDF tente donc de réaliser tout le potentiel du document électronique. Les qualités de ce format lui assurent un succès important, et il faut compter que celui-ci ira en grandissant.

Les formats PostScript et PDF sont tous deux gérés par la société Adobe, qui précise dans le *PostScript Language Reference, 3rd edition* que la société détient le droit d'auteur sur la spécification écrite du langage, mais que toute personne peut 1° Écrire un programme dans le langage PostScript ; 2° Écrire des programmes qui génère comme sortie un programme se conformant à la spécification PostScript ; 3° Écrire des logiciels qui interprètent des programmes PostScript. ¹⁴ La

14. Ceci fut initialement l'avantage commercial d'Adobe, qui retirait des droits pour chaque imprimante laser qui incorporait un interprète PostScript, fourni exclusivement par Adobe.

norme n'est donc pas ouverte, mais ceci ne l'a pas empêché d'être un instrument fédérateur de l'industrie, et n'a pas nécessairement nui à son adoption.

2.7.1 Exemple d'un fichier PostScript

```
%!PS-Adobe-3.0
%%Title: (aae.doc)
%%CreationDate: (11:43 Lundi 16 octobre 2000 )
%%Pages: 1
%%Orientation: Portrait
%%EndComments
[...]
%%BeginFeature: *PageSize A4Small
<</PageSize [595 842] /ImagingBBox null>> setpagedevice
%%EndFeature
%%EndSetup
%%Page: 1 1
[...]
gS 0 0 538 781 rC
86 75 :M
f57 sf
-.174(Longue vie \210 l\325acte)A
158 123 :M
-.192(authentique)A
143 171 :M
-.181(\216lectronique!)A
endp
showpage
%%Trailer
end
%%EOF
```

Le fichier commence par un certain nombre de commentaires, ignorés par l'interprète PostScript, qui fournissent certaines informations sur les conditions de création du fichier — titre, nombre de pages, etc. Ensuite, la commande `BeginFeature` montre une commande envoyée à l'interprète indiquant que le document est de format A4. La troisième partie du fichier montre les commandes pour afficher le texte (les accents sont représentés par leur code ASCII). La dernière commande, `showpage`, enjoint à l'interprète d'imprimer la page.

2.8 Pourquoi une norme?

A l'issue de ce rapide survol, on voit qu'il n'existe pas de solution évidente à la question de l'encodage des actes authentiques — aucun des formats d'encodage proposés n'a été conçu pour répondre aux exigences spécifiques de ce type de document. Le décret doit-il alors imposer l'adoption d'une format d'encodage précis pour l'acte authentique électronique? Si c'est le cas, quelles devraient être les caractéristiques de cette norme, et pourquoi? Plus problématique encore, si chacun des formats d'encodage fait jouer une machinerie informatique complexe dont le résultat final est la visualisation d'un document, à l'écran ou sur papier, ce processus n'a que peu à voir avec le concept de « lisibilité » tel qu'on l'entend dans le contexte de l'écrit papier. Comment doit-on

repenser ce concept de lisibilité, étant donné la nouvelle donne technologique du document ?

Certains ont suggéré qu'il était essentiel que le format adopté soit 1°) normalisé ; 2°) que cette norme soit stabilisée ; 3°) qu'elle soit aussi ouverte que possible, de façon à favoriser les échanges ; 4°), que ce format soit transparent, de façon à garantir au signataire qu'il n'existe aucune donnée ou instruction cachées dans le fichier qu'il valide. Il est utile d'examiner attentivement chacun de ces arguments, car ils reviennent régulièrement dans les discussions portant sur les normes.

Norme : Tout d'abord, pourquoi faut-il déterminer un format particulier d'encodage pour l'acte authentique ? Qu'y gagne-t-on, mais aussi, qu'y risque-t-on ? Est-ce que d'adopter une norme est la réponse aux besoins exprimés par les acteurs de l'acte authentique ? Les normes de format de fichier décrits dans ce chapitre sont, avant toute chose, des instruments stratégiques industriels : leur évolution est soumise à des impératifs qui ont peu à voir avec la logique de l'acte authentique.

Stabilité : Il est difficile de jauger précisément ce que représente la stabilité d'une norme technique, dans l'univers informatique. Parmi les formats que nous avons envisagé dans cette section, le XML est certainement la moins stabilisée de toutes, mais son ampleur probable procure un certain gage de stabilité future.

Ouverture : On entend souvent vanter les mérites d'une norme, en ce qu'elle est « ouverte », par opposition à une autre qui serait, plus vénalement, « propriétaire ». Une norme ouverte est une norme dont le contenu est non seulement connu de plus, mais en plus, qui est déterminée par un processus de discussion et de consultation entre les acteurs du domaine d'intérêt de la norme. Un format propriétaire est défini par une seule société — par exemple, le format de fichier du traitement de texte Word 2000 ne dépend que de la société Microsoft. Malheureusement, la plupart des normes ne peuvent être identifiées à un de ces deux pôles, mais se situent plutôt quelque part dans une vaste zone intermédiaire. En fait, les nouvelles institutions à la base du développement des standards Internet sont difficiles à aligner selon une grille classique « institutions publique \longleftrightarrow institutions privées » qui produiraient des normes respectivement « ouvertes \longleftrightarrow propriétaires » : l'IETF, le W3C, l'ICANN sont des instances de normalisation encore difficiles à classer.

Transparence : L'argument revient souvent qu'il faut éviter qu'un fichier informatique encodant un acte authentique électronique ne puisse contenir un virus ou une macro susceptible de modifier le contenu de l'acte. Le virus « I LOVE YOU » n'était-il pas précisément une telle instance d'un « document exécutable », capable de faire exécuter par l'ordinateur toute une série d'actions néfastes ? Selon cet argument, un format d'encodage « transparent », non susceptible de contenir des instructions informatiques serait plus sécuritaire qu'un format « non-transparent ». Une telle analyse ne tient cependant pas compte du fait qu'un document électronique est le résultat de l'interaction entre des composantes fichiers, des composantes logicielles, et des composantes matérielles : un document « transparent » comme du HTML ou du XML peut tout à fait faire appel à des ressources (code Java par exemple) distribuées ailleurs sur le réseau, susceptible d'avoir une influence sur le document. En fait, le problème est beaucoup plus étendu que la simple possibilité de virus cachés dans les fichiers : sur un ordinateur commercial, il est à toute fin pratique

impossible de *garantir* que le processus de signature ne soit usurpé, à un moment ou un autre.¹⁵ La notion du *What You See Is What You Sign* ne capture qu'une petite partie de ce problème et va exiger d'être considérablement raffinée et explicitée avant qu'elle puisse être opérationnalisée, tant juridiquement que techniquement.

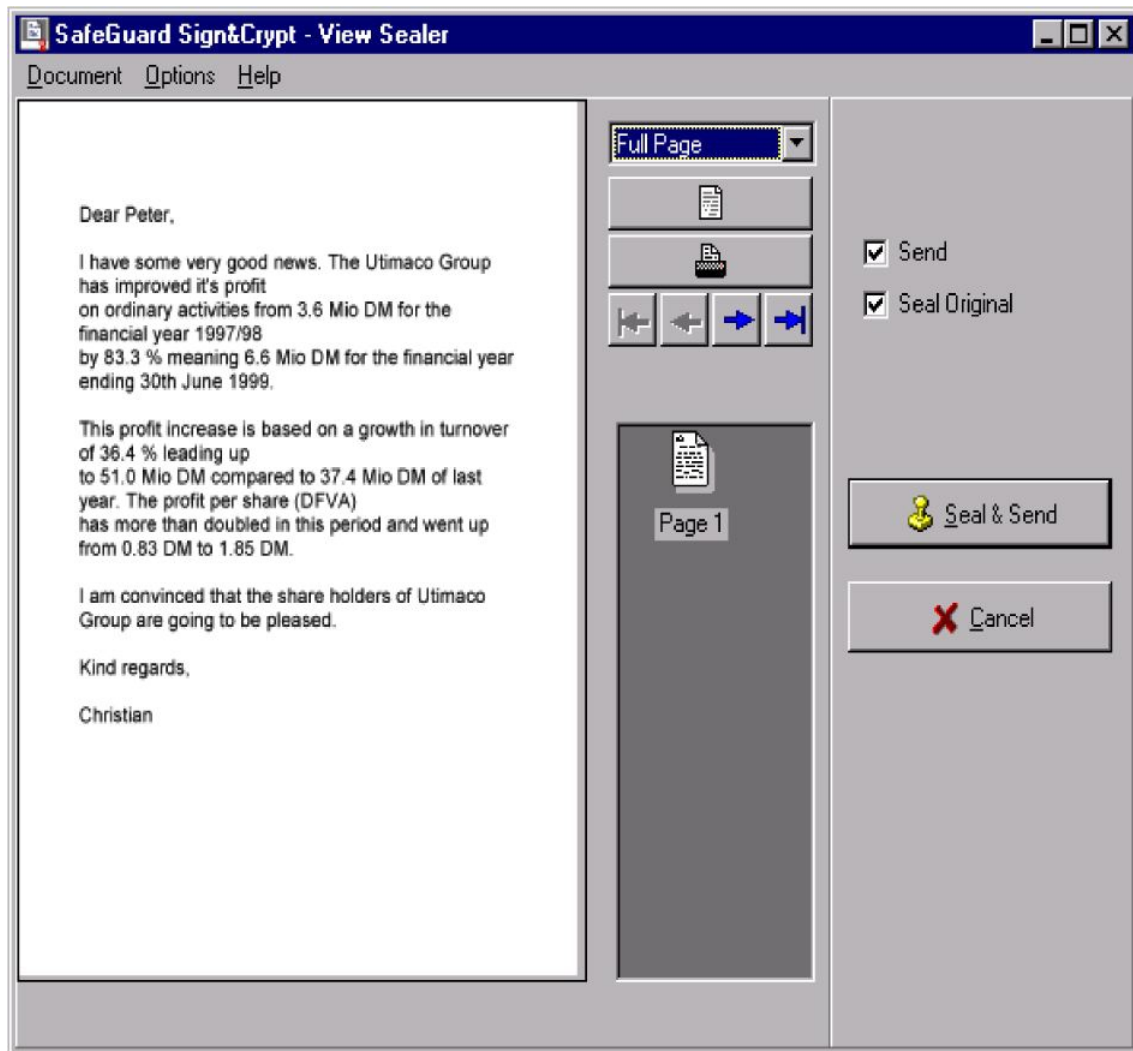


FIG. 2.1 – Utilisation du format TIFF pour le What You See Is What You Sign

Quelque soit la validité de cette analyse, le format TIFF a été adopté pour la réalisation de procédés de signature électronique réalisant le « *What You See Is What You Sign* », par exemple, le procédé « Sign&Crypt » de la société Utimaco : l'idée est d'utiliser un pilote d'impression qui traduit le fichier à signer (fichier Word, par exemple) en un fichier TIFF, de l'afficher à l'écran et de signer ce fichier — voir la figure 2.1. On peut envoyer les deux fichiers, le fichier original et

15. Voir à ce sujet l'analyse de Bruce SCHNEIER, qui considère que la seule solution réside dans l'utilisation d'un terminal de signature dédié et certifié, un ordinateur personnel multi-usages étant, à tout fins pratiques, impossible à sécuriser. Voir Bruce SCHNEIER *Secret and Lies: Digital Security in a Networked World*, John Wiley and Sons, 2000.

sa version TIFF signée, au destinataire, de façon à pouvoir continuer à travailler sur le document original, tout en disposant d'une image signée du document.

2.9 Quel format d'encodage pour l'acte authentique électronique?

Si l'on décide d'imposer un format d'encodage à l'acte authentique électronique, quel devrait en être les caractéristiques? Les premières réflexions ont désigné le format XML comme celui présentant le plus d'avantages, et ce, pour un certain nombre de raisons : 1°) il permet de distinguer le fond et la forme ; 2°) il se présente sous la forme d'un texte brut encadré par des balises ordonnées hiérarchiquement suivant les spécifications de DTD ou un schéma XML ; 3°) il est directement lisible par l'homme au prix d'un certain effort ; 4°) sa représentation est très simple dans la mesure où un document XML contient uniquement un texte et des balises à l'exclusion de toute macro-commande dans laquelle pourrait être dissimulée un virus ou des données cachées qui viendrait modifier le document après archivage ; 5°) la vision à l'écran ou à l'impression du texte sera toujours la même quel que soit le poste de consultation ou le type de plate-forme ; 6°) il peut être facilement obtenu à partir d'un traitement de texte standard. Il peut être instructif d'analyser ces arguments de plus près.

Le XML permet effectivement de distinguer la forme du fond, mais, dans le contexte de l'acte authentique, c'est plutôt un désavantage, car l'acte authentique, justement, lie intimement la forme et le fond!¹⁶ Le second argument n'est qu'un constat ; Pour le troisième, pour qu'un document XML soit « directement lisible » par l'homme, il faudrait qu'il puisse simultanément garder à l'esprit le fichier XML, le fichier DDT, et la feuille de style XSL associée, ainsi que tous les documents externes possiblement référencés ; le quatrième argument a été discuté plus haut ; le cinquième est extraordinaire, puisqu'il est difficile d'imaginer comment on peut à la fois dissocier la forme du fond *et*, simultanément, assurer qu'un document aura la même apparence quel que soit le périphérique de visualisation ! ; sixièmement, *chacun* des formats discutés dans cette section sont facilement obtenus à partir d'un traitement de texte standard (Word, par exemple) — texte « brut », RTF, HTML, XML, PostScript, PDF, TIFF.

Il faut donc reposer la question : Quels doivent être les critères de sélection d'un format d'encodage, étant donné les particularités de l'acte authentique électronique ? Comment traduire les concepts de lisibilité et de pérennité au niveau du document électronique ?¹⁷

Il faut d'abord considérer que d'autres logiques économiques et techniques sont à l'œuvre. Au niveau de la rédaction de l'acte, le passage à L'XML ne fera que rationaliser un processus déjà largement entamé : la plupart des logiciels de rédaction d'actes combinent actuellement différentes bases de données (bases de clients, clausier, etc) pour produire un modèle de document qui est ensuite transféré à Word pour être sauvegardé et imprimé. L'information est donc déjà structurée,

16. Voir à ce sujet les remarques de Mme Yvonne FLOUR, « Les conventions sur la forme » *Répertoire Desfresnois*, 15-16/00, p.911-928, section II: « La forme définit le fond ».

17. Plusieurs excellents documents de réflexion ont été produits à ce sujet par le *Council on Library and Information Resources* — voir *Authenticity in a Digital Environment*, Council on Library and Information Resources, mai 2000 et G. Lawrence, W. Kehoe, O. Rieger, W. Walters, et Anne Kenney, *Risk management of digital information: a file format investigation*, Council on Library and Information Resources, juin 2000, disponibles sur www.clir.org.

mais elle n'est pas susceptible d'être échangée, puisque chaque logiciel définit son propre format de données. Le XML permettra de standardiser ce processus, qui aura alors l'avantage de pouvoir permettre l'échange de données directement entre les administrations et d'automatiser les étapes qui peuvent l'être. Pour tout ce qui concerne les aspects d'échanges de données, le passage à l'XML semble donc assez logique, puisque celui-ci semble vouloir s'imposer comme norme globale de structuration de l'information. La stabilisation de cette norme (et des autres qui lui sont afférentes) sera sans doute néanmoins relativement longue, car on tente ici de définir le langage qui permettra d'effectuer *tous les types d'échanges de toutes les sortes d'informations* par le Web.

Le problème, c'est que le XML ne fixe que faiblement la forme visuelle du document — il faut disposer des trois éléments du document, XML, DDT et XSL, et des programmes informatiques qui ont assemblé ces trois fichiers en la représentation visuelle d'un acte juridique. Prétendre que le XML est lisible « directement » ne résout pas la question, puisque c'est faire l'économie de la réflexion qui est l'objet même de la constitution de ce groupe de travail : Que devient la notion de lisibilité à l'ère informatique ? Comment s'assurer que cette lisibilité demeure pérenne ? Il est peu probable que le critère de la « lisibilité directe par l'homme » du fichier d'encodage du document soit la réponse à cette question.

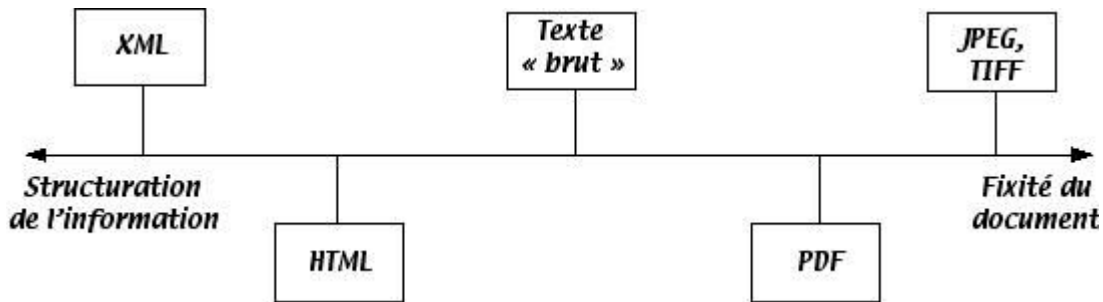


FIG. 2.2 – Comparaison des différents formats d'encodage

Les formats d'encodage peuvent être classifiés selon un axe qui oppose structuration de l'information à fixité de la représentation (voir la figure 2.2) : on peut en conclure que l'utilisation du XML comme format d'encodage des actes authentiques électronique est appropriée pour toutes les phases de traitement de l'acte où la finalité d'échange d'information prime, alors que pour les phases de l'acte où l'intégrité visuelle du document prime, il est préférable de se porter vers des formats d'encodage qui nécessitent une machinerie moins lourde que le XML et qui ne concernent que la représentation visuelle de l'acte (PostScript, PDF, JPEG et TIFF).

Chapitre 3

Signature

Plusieurs difficultés se présentent avant même que ne soit possible une discussion de la signature électronique : tout d’abord, le vocabulaire entourant la signature électronique est tributaire des guerres rangées entre les intérêts commerciaux qui s’affrontent pour la conquête des marchés naissants entourant la sécurisation des transactions. Ainsi, alors qu’initialement on utilisait le terme de « signature électronique » pour désigner l’ensemble des technologies de signature et le terme de « signature numérique » pour désigner la signature basée sur la cryptographie, plusieurs documents récents font l’adéquation entre signature électronique et signature cryptographique — par exemple, les rapports de M. Jolibois au Sénat et de M. Paul à l’Assemblée nationale.¹ Dans ce rapport, on utilisera le terme « signature électronique » pour désigner une technologie de signature qui est, à un moment ou à un autre du cycle de vie d’un document, électronique, sans même exclure qu’elle existe sous une forme non-électronique à d’autres moments — ceci dans la perspective d’un œcunémisme technologique le plus inclusif possible.

Ensuite, il faut replacer la signature dans un *contexte global* de sécurité des documents : dans l’immense majorité des cas, l’authenticité — pas celle du droit français, mais la notion plus informelle — du document ne sera jamais contestée, et même dans le cas contraire, la validité de la signature ne sera pas nécessairement au centre du débat, comme le souligne Benjamin Wright :

« Malgré ce que l’on pourrait en déduire à l’écoute des dramatiques de télévisions axés sur l’univers judiciaire et le procès, l’origine et l’authenticité de la plupart des documents n’est pas prouvée sur la base de signatures. Beaucoup plus souvent, l’origine et l’authenticité sont prouvés à partir d’un ensemble de faits et de circonstances — l’ensemble des relations entre les gens impliqués — le moment, le lieu, les transactions et discussions afférentes, la fonction et le contenu des documents, et non simplement la signature. Plus souvent qu’autrement, l’origine et l’authenticité des documents n’est pas disputée et les parties déclarent simplement que les documents sont bien ce qu’ils apparaissent être. »²

Ces remarques sont valables tant pour l’univers du papier que pour l’univers électronique : dans un cas comme dans l’autre, la signature n’est qu’un seul maillon d’une chaîne complexe de fac-

1. De même pour de nombreux commentateurs, par exemple, Thierry PIETTE-COUDOL : « dans la signature électronique, le certificateur vient garantir la clé publique (cryptographique) que le destinataire d’un message signé utilisera pour vérifier la signature. » — *Échanges électroniques, Certification et sécurité*, Paris : Litec, 2000.

2. Benjamin WRIGHT, “The Legality of the PenOp Signature,” PenOp White Paper, disponible sur le site www.penop.com.

teurs qui, ensemble, garantissent l'authenticité des documents.

Quatre technologies sont ici discutées, chacune représentant trois approches très différentes au problème de la signature électronique : la signature cryptographique,³ la signature biométrique, la signature-tatouage et la signature numérisée. Nous adoptons une attitude agnostique quant à la supériorité de ces méthodes les unes par rapport aux autres.⁴ Il faut malheureusement considérer de tels arguments à la lumière des énormes enjeux commerciaux qui entourent l'adoption de ces technologies. Nous considérons que chacune de ces technologies présente des avantages et des inconvénients, mais que ces ceux-ci et ceux-là ne peuvent être évalués qu'à partir d'une caractérisation précise du contexte d'utilisation envisagé.

A tout seigneur, tout honneur : puisque pour beaucoup, la signature électronique est, par définition, la signature cryptographique, c'est cette technologie que nous commencerons par ... déchiffrer.

3.1 La signature cryptographique

Si l'histoire de la cryptographie peut être retracée, selon David KAHN, aux premiers hiéroglyphes hors-normes,⁵ elle a connu un essor considérable depuis le début des années 70, alors que son utilité pour sécuriser les échanges bancaires a été reconnu. C'est à cette occasion que la recherche fondamentale en ce domaine devait commencer à se dégager du cadre strictement militaire qui était jusqu'alors le sien et connaître sa première heure de gloire, lorsque Whitfield Diffie et Martin Hellman conçoivent en 1976 le principe de la cryptographie à clé publique et suggèrent pour la première fois que cet outil pourrait fournir un équivalent à la signature au sein des environnements électroniques.⁶

Tout d'abord quelques notions de base : jusqu'à l'invention de Diffie et Hellman, la cryptographie était basée sur le principe suivant, dit de *cryptographie symétrique*, parce que les participants utilisent la même clé tant pour le chiffrement que pour le déchiffrement. Soit deux individus, A et B , désirant communiquer au sein d'un environnement potentiellement hostile. A dispose d'un procédé de chiffrement C et B dispose du procédé de déchiffrement inverse D . La confidentialité de leur communication repose sur une *clé secrète* K , connue à la fois de A et B , sur laquelle ils se sont entendu au préalable. Pour chiffrer un document M , A fournit son document M et la clé K au procédé de chiffrement E , qui produit un document chiffré M' . Ce document est transmis à B par un canal de transmission quelconque, avec l'hypothèse qu'il est impossible à quiconque ne dispose pas de la clé K de déchiffrer M' . Après avoir reçu le document M' , B le fournit, ainsi que

3. Les termes de *signature numérique* ou de *signature digitale* sont aussi souvent utilisé pour désigner la signature cryptographique.

4. Cette attitude n'est pas le fruit d'un quelconque relativisme technologique : même quand il s'agit de comparer la sécurité de différents modèles de cryptographie asymétrique — celle-ci peut en effet être réalisée à partir de différents problèmes mathématiques — aucun spécialiste ne s'engage, faute de grille qui permettrait de les comparer. Quand il le fait, il s'agit alors d'un acte de foi ou, plus pragmatiquement, de simple rhétorique commerciale.

5. David KAHN, *La guerre des codes secrets : Des hiéroglyphes à l'ordinateur*, Inter Éditions, 1980 ; un autre ouvrage populaire sur l'histoire de la cryptographie est celui de Simon SINGH, *Histoire des codes secrets : De l'Égypte des pharaons à l'ordinateur quantique*, Lattès 1999 ; un excellent exposé vulgarisé de la cryptographie moderne est celui de Jacques STERN, *La science du secret*, Odile Jacob 1998.

6. W. DIFFIE, M. E. HELLMAN, « New Directions in Cryptography », *IEEE Transactions on Information Theory*, IT-22, pp. 644-654, novembre 1976.

la clé K , au procédé de déchiffrement D , qui produit alors le document M original, en clair.

Diffie et Hellman s'intéressaient au problème de la sécurisation *efficace* des communications entre n individus présents sur un réseau de communication, le réseau téléphonique par exemple. En utilisant les méthodes de cryptographies à clé symétrique, deux scénarios sont possibles : soit que chaque paire d'individus établit une clé de chiffrement commune, dont chacun possède un exemplaire ; soit qu'une autorité centrale se charge de coordonner l'échange de clés entre individus. Dans le premier cas, il est nécessaire de gérer n^2 clés sur le réseau, par le biais de n^2 communications ; dans le second, on se doit d'investir de confiance une tierce partie, chose toujours regrettable, du point de vue d'un cryptologue (voir mes remarques plus bas). Diffie et Hellman ont voulu éviter ces deux inconvénients d'un seul coup.

La *cryptographie à clé publique* (ou asymétrique) procède d'un principe simple, mais très astucieux. A chaque individu présent au sein du réseau de communication, est attribuée une paire de clés, une *clé publique* et une *clé privée* (on parle parfois de bi-clés), qui permettent de réaliser et le chiffrement, et la signature. La clé publique de chaque individu est rendue disponible au sein d'un annuaire, alors que la clé privée est conservée secrète. Pour envoyer un message M chiffré à B , A obtient la clé publique de B et s'en sert pour chiffrer le message. De son côté, B est en mesure de déchiffrer le message en utilisant sa clé privée. Pour signer un message, A utilise sa clé privée avant de transmettre le message à B . Celui-ci est en mesure de vérifier l'identité de l'expéditeur en se procurant la clé publique de A .⁷ Toute la magie du système repose dans l'hypothèse mathématique suivante : *bien que la clé publique et la clé privée soit uniquement complémentaire, même en connaissant la clé publique, il est impossible d'en déduire la clé privée.*⁸

Les conséquences pratiques d'un tel procédé sont importantes : tout d'abord, il n'est plus nécessaire à deux individus, désireux d'échanger des données signées et/ou chiffrées au sein du réseau, de s'entendre au préalable sur une clé, chaque individu se contentant de publier sa clé publique dans un répertoire ; ensuite, le nombre de clés nécessaires sur le réseau passe de l'ordre de n^2 à $2n$, une conséquence d'une grande importance pratique — pour $n = 1000$, on passe d'un million de clés à gérer à 1000 bi-clés. Et, Diffie et Hellman y ont vu l'opportunité de décentraliser la communication des clés, c'est-à-dire d'éviter de faire appel à une autorité de confiance — il faut se replacer dans le contexte post-Watergate de l'époque, de la méfiance caractérisée des citoyens américains face à l'intrusion de l'État dans leur vies, et de l'énorme emphase mise sur la liberté d'expression au sein de cette société.

Initialement, Diffie et Hellman n'ont pu concevoir une réalisation concrète de leur principe et on du se contenter d'en énoncer les principes. Ce n'est que deux années plus tard que Ronald Ri-

7. En pratique, pour éviter de signer de longs messages, les systèmes de signature font appel à des *fonctions cryptographiques de hachage*, qui produisent tout d'abord un *condensé* du message. Un condensé est une représentation du message de taille fixe, avec la particularité mathématique qu'il est « impossible » à un fraudeur de déterminer deux messages qui produirait un condensé identique. C'est le condensé qui est signé, par souci d'efficacité.

8. Il importe de préciser que ceci demeure un hypothèse, c'est-à-dire qu'aucun procédé de cryptographie à clé publique ne peut fonder sa sécurité sur une preuve mathématique : elle est plutôt fondée sur des *hypothèses calculatoires*. Un seul système de chiffrement (symétrique) peut se vanter d'une sécurité absolue, le système dit de *one-time pad*. Cependant, les conditions pratiques de son utilisation (clé de même taille que le message et exigence de renouveler la clé *pour chaque message*) rendent son utilisation confinée aux applications aux plus hautes exigences de sécurité.

vest, Adi Shamir, et Leonard Adleman, du Massachusetts Institute of Technology allaient concevoir d'un procédé mathématique qui mette en œuvre le principe de la cryptographie à clé publique, procédé fondé sur l'opération mathématique de l'exponentiation modulaire.⁹ C'est ce processus qui est le plus largement répandu aujourd'hui.¹⁰

3.1.1 La certification

La solution de Diffie et Hellman souffrait cependant d'une faiblesse importante, propre à invalider les bénéfices du système : supposons qu' \mathcal{O} , un être fourbe et malhonnête, désire convaincre \mathcal{A} qu'il reçoit des messages signés de \mathcal{B} , alors qu'ils sont en fait de la plume d' \mathcal{O} . Celui-ci n'aurait qu'à substituer sa propre clé publique à celle de \mathcal{B} dans l'annuaire, et envoyer ses messages à \mathcal{A} en prétendant être \mathcal{B} . Pour vérifier la signature de ces messages, \mathcal{A} se procurerait la clé publique de \mathcal{B} (en fait, celle d' \mathcal{O}) et la vérification étant réussie, serait faussement convaincu de l'origine des messages. Il faut donc que les clés publiques soient obtenues de telle façon à ce que l'on soit convaincu de l'identité de la personne reliée à la clé. Deux méthodes ont été explorées pour résoudre ce problème : la première, dite des *réseaux de confiance*, est basée sur la transmission des clés entre les individus eux-mêmes. Lorsque vous recevez une clé publique, vous évaluez informellement son authenticité, et la transmettez à vos propres correspondants. Chaque clé bénéficie en quelque sorte d'une évaluation de sa qualité, selon la confiance que vous portez en la personne qui vous l'a transmise. C'est cette méthode qui est utilisée dans les versions gratuites du logiciel PGP (mais non dans la version commerciale, qui est basée sur la certification).

Une seconde méthode a été imaginée, présentant le désavantage de reproduire le besoin pour une autorité centralisée au sein du réseau. Elle consiste en l'utilisation d'un *certificat* et d'une *autorité de certification*. Un certificat est tout simplement un document contenant une série d'informations permettant d'associer une clé publique à un individu. Par exemple, un certificat répondant aux exigences de l'annexe I de la Directive européenne sur la signature électronique pourrait res-

9. Cette opération possède la particularité que si l'on connaît des algorithmes efficaces qui permette de l'effectuer, on n'en connaît pas pour *renverser* l'opération, c'est-à-dire retrouver les données de départ de l'opération en partant du résultat. Ceci est plus facile à voir avec l'exemple de la multiplication : tout nombre possède une décomposition unique en nombre premiers — $15 = 3 \times 5$; $16 = 2 \times 2 \times 2 \times 2$. On sait comment efficacement multiplier 3 et 5 ou 4 et 4. Par contre, on ne connaît pas d'algorithme efficace pour effectuer l'opération inverse, déterminer les facteurs premiers d'un nombre. Pour un très grand nombre, cette inefficacité est si coûteuse qu'on dit, par abus de langage, qu'elle est, à toutes fins pratiques, impossible, c'est-à-dire qu'elle nécessiterait des milliards d'années de calcul aux ordinateurs les plus puissants.

10. Le fait que l'algorithme RSA soit utilisé dans le mécanisme de sécurisation des transactions SSL des fureteurs Web a fait dire à son fabriquant que l'algorithme RSA est le bout de code informatique aujourd'hui le plus répandu dans le monde.

sembler à :

```
-----  
Nom          :   BLANCHETTE  
Prénom       :   Jean-François  
Clé publique :   AD3456EBE12976EDE  
Emission     :   01/01/2000  
Expiration   :   31/12/2001  
Limite de valeur : 1000 euros  
Emis par     :   CNRS  
Algorithme   :   DSA 1.78  
Numéro de série : 34343343434343433  
-----
```

La véracité des informations contenues dans le certificat est confirmée par deux processus distincts : 1° d'une part, l'autorité de certification engage une procédure par laquelle l'identité de la personne est confirmée — présentation en personne de pièces d'identités, etc. ; 2° d'autre part, le certificat est lui-même signé électroniquement par la clé privée de l'autorité de certification. Toute personne qui désire vérifier la véracité du lien entre un individu et sa clé publique peut dorénavant le faire en 1° vérifiant la signature de l'autorité de certification sur le certificat de l'individu, en se procurant la clé publique de l'autorité ; 2° se convaincant, par tous les moyens à sa disposition, de la probité de l'autorité de certification.

D'entrée de jeu, plusieurs questions se posent : *qui va certifier la clé publique des autorités de certification ?* Cette question fort complexe est à l'origine de plusieurs modèles, et devrait, à elle seule, donner naissance à une riche industrie juridique dans les prochaines années : certification croisée, auto-certification, certification hiérarchique, autant de modèles différents qui induisent des modes de vérification de signature différentes. Le point le plus important à retenir est que la vérification d'une signature implique la vérification de la totalité de la *chaîne de certification*, c'est-à-dire l'ensemble des certificats des autorités de certification impliquées.

Un individu qui veut, aujourd'hui, obtenir un certificat à clé publique, peut le faire de différentes façons. Soit qu'il l'obtient d'une autorité de certification à partir du réseau. Ce certificat peut ensuite être intégré directement à son ordinateur (Windows 2000 supporte les certificats), soit que ce certificat est intégré à son logiciel de courriel, ou encore à son fureteur Web, qui intègrent différentes fonctions de gestion de certificats.

3.1.2 Les infrastructures à clés publiques

On utilise le terme *infrastructure à clés publiques* (ICP)¹¹ pour désigner la combinaison d'éléments matériels, logiciels, et procéduraux qui permette d'effectuer l'ensemble des opérations sous-jacentes à la réalisation de la signature (et du chiffrement) cryptographique, c'est-à-dire :

- 1° *Génération de clés cryptographiques* : il faut produire les paires de clés cryptographiques, de façon hautement sécuritaire ;
- 2° *Distribution des clés privés aux utilisateurs* : il faut que les clés privés soient distribuées aux utilisateurs — placées au sein d'une carte à puce, par exemple ;

11. Ou encore, PKI pour *Public-Key Infrastructures*, ou IGC pour *Infrastructure de gestion de clés*.

- 3° *Enregistrement des utilisateurs* : Il faut que l'autorité de certification vérifie l'identité de chacun des utilisateurs, par exemple par une présentation en personne de pièces d'identités ;
- 4° *Certification des clés publiques* : Une fois l'identité des utilisateurs confirmée, l'autorité de certification doit rédiger le certificat et le signer avec sa clé privée ;
- 5° *Service d'annuaire* : Le certificat doit ensuite être placé au sein d'un annuaire, de façon à permettre à d'autres utilisateurs de vérifier les signatures ;
- 6° *Révocation des certificats compromis ou périmés* : Les certificats sont révoqués lorsque ils sont expirés, ou lorsque une clé privée a été compromise. Ainsi, une signature ne pourra être vérifiée à l'aide d'une clé publique dont le certificat est révoqué ;
- 7° *Archivage* : Il faut conserver l'ensemble des certificats qui permettent la vérification, les listes de révocation, etc., de façon à pouvoir effectuer la vérification ultérieurement ;
- 8° *Recouvrement des clés* : Dans plusieurs pays, notamment la France, la Grande-Bretagne, et les Etats-Unis, les forces de l'ordre ont exprimé de grandes réserves face à l'impossibilité de pouvoir déchiffrer des messages qui circulent sur les réseaux. Les technologies de recouvrement de clés permettent, de différentes façons, d'accéder aux clés de chiffrement d'un utilisateur, ou encore, de récupérer les clés de déchiffrement si elles étaient égarées ;
- 9° *Horodatage* : les certificats et les signatures doivent faire l'objet d'un datage sûr.

Une organisation désirant déployer une PKI peut déléguer l'ensemble de ces fonctions à un prestataire, ou au contraire, les réaliser toutes ou en partie. Cette liste permet de constater d'un coup d'œil la complexité de l'infrastructure sous-jacente à la signature cryptographique.¹² Et l'on ne décrit même pas ici les difficultés liées au déploiement, aux facteurs organisationnels et culturels. Nous aurons l'occasion d'examiner de plus près le déploiement et le fonctionnement d'une PKI dans le chapitre 5, portant sur l'archivage.

3.1.3 Brèves remarques sur l'histoire de la cryptographie

Effectuons un bref retour sur les circonstances historiques du développement de la science cryptographique. Les technologies de chiffrement ont été conçues dans un contexte militaire, pour permettre à des individus ou des institutions de communiquer au sein d'un environnement *ouvert et hostile*. En cryptographie, un système robuste est un système qui fonctionne au sein de l'environnement le plus hostile imaginable, un environnement où les intervenants malhonnêtes tenteront de faire échouer l'objectif de sécurité par tous les moyens possibles.

Prenons l'exemple des premières techniques du chiffrement, qui faisaient reposer leur sécurité en partie sur le fait que l'ennemi ignorait la technique utilisée, le secret reposant dans l'indisponibilité de la technique en quelque sorte. C'est le Français Auguste Kerckhoffs qui, le premier, a suggéré qu'il était préférable de poser l'hypothèse que l'ennemi avait pleine connaissance de la technique utilisée et que le secret devait reposer dans une unité d'information facilement modifiable et renouvelable : *la clé*.¹³ La progression est ici d'une hypothèse de méfiance faible (« l'ennemi ne connaît

12. Face à la lourdeur et la complexité des PKIs, de nouveaux modèles se sont développés, qui raffinent le concept de certification : les LPKIs, pour *Lightweight PKI*. Au sein de communautés d'utilisateurs réduites, ce concept est promoteur. Voir à ce sujet le système Securicam de la société Atos (www.atos-group.com.)

13. Kerckhoffs énonce comme deuxième exigence d'un système de chiffrement militaire qu'« il faut qu'il n'exige pas le secret et qu'il puisse sans inconvénient tomber entre les mains des ennemis ; » — Auguste Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. IX, pp. 5–38, Janvier 1883, pp. 161–191, Février 1883.

pas ma méthode de chiffrement »), à une hypothèse de méfiance plus forte (« l'ennemi a pleine connaissance de ma méthode de chiffrement »). La recherche en cryptographie progresse de cette façon : votre système est meilleur s'il peut résister à des adversaires plus puissants et plus hostiles que le système précédent ne le pouvait. En d'autres termes, il est meilleur s'il vous permet d'investir moins de confiance en votre environnement — et de la faire plutôt reposer sur les mérites de votre système cryptographique.

C'est ainsi un des objectifs fondamentaux de la recherche en cryptographie que de faire reposer la sécurité d'un système sur les hypothèses de confiance les plus faibles possibles. Ceci est un point important, puisqu'il offre une clé de lecture fondamentale des technologies cryptographiques : *elles tendent à vouloir réduire le plus possible la confiance que l'on doit accorder aux intervenants institutionnels*. Le très réputé cryptologue David Chaum résume bien cette façon de penser :

« Ce que nous avons pu démontrer à l'aide des technologies modernes de l'information, c'est que l'on peut résoudre tout problème de sécurité de l'information simplement en laissant chaque personne posséder son propre ordinateur. Laissez-les utiliser leur propre ordinateur pour protéger leurs propres intérêts. *Il n'y a aucun besoin d'établir des mécanismes auxquels les parties accordent mutuellement leur confiance.*¹⁴»

Ainsi, les technologies cryptographiques ne sont pas neutres, bien au contraire, puisqu'elles tendent à nier la composante institutionnelle de la sécurité de l'information. Ainsi, face à un discours technique qui privilégie systématiquement une approche algorithmique aux différents problèmes de la sécurité de l'information, il faut demeurer vigilant et privilégier une approche qui harmonise les technologies aux institutions concernées.

3.1.4 Normalisation

Au niveau de la normalisation, l'univers cryptographique est très avancé.¹⁵ Très tôt, les industriels et les scientifiques ont compris que le succès de cette technologie dépendrait de leur capacité à développer des normes universellement acceptées. De plus, la sécurité informatique s'est toujours bien prêtée à la normalisation — l'administration américaine a stimulé l'essor de la cryptographie civile en établissant les *Federal Information Processing Standards* (FIPS) qui définissent, entre autres, le procédé de chiffrement symétrique DES et la norme de signature digitale DSA. L'*Internet Engineering Task Force* travaille sur un certain nombre de normes — fonction de condensés (MD5) et courrier électronique sécurisé entre autres (S/MIME), PKIs, etc. Certains standards se sont imposés *de facto*, c'est le cas de la série de spécifications PKCS (*Public-Key Cryptographic Standards*) où est définie la norme de chiffrement RSA et l'échange de clé Diffie-Hellman. Ces spécifications ont joué un rôle important dans l'opérationnalisation des systèmes cryptographiques.

La communauté académique a pu imposer l'idée que l'échange et la discussion ouverte qui caractérise le processus scientifique seuls permettraient de garantir que les procédés cryptographiques soient sûrs. Tout procédé propriétaire, qui ne puisse être soumis à l'examen de l'ensemble

14. David Chaum, "Digital Money." Présentation donnée à la conférence *Doors of Perception 2*, Amsterdam, 4-6 novembre 1994.

15. Voir à ce sujet Menezes, van Oorschot et Vanstone, *Handbook of applied cryptography*, CRC Press, 1997, chapitre 15.

de la communauté scientifique, serait susceptible de contenir des faiblesses non-anticipées par ses concepteurs. Cette vision résultante de l'expérience de la communauté scientifique de la difficulté de réaliser des procédés cryptographiques qui réalisent leurs objectifs de sécurité et que l'ensemble des conditions à considérer est extrêmement vaste et complexe.

Il ne faut cependant pas glorifier le processus scientifique comme garant d'une sécurité absolue. Chaque année, les jeunes cryptologues se font les dents en brisant des procédés cryptographiques dont la sécurité a pourtant été *mathématiquement prouvée*. C'est que, contrairement à la vision populaire de la chose, la preuve mathématique n'est pas unique, elle existe sous différentes formes et il faut donc au préalable que les scientifiques s'entendent sur ce que constitue une preuve valable. Encore plus important, il faut qu'ils s'entendent sur les définitions des objets dont ils tentent de prouver les propriétés. Or, pour beaucoup d'objets cryptographiques, ces définitions ne sont pas encore stabilisées.¹⁶

D'un côté, donc, le processus d'examen scientifique ouvert des procédés cryptographiques est probablement le mieux à même de produire des technologies dont on soit raisonnablement assuré de la fiabilité. D'un autre côté, il n'est pas clair quelle soit l'importance de cette fiabilité mathématique dans la chaîne de sécurité. Contrairement au discours dominant, il est fort possible qu'elle ne joue, somme toute, qu'un rôle très mineur.

3.1.5 Mesure de la sécurité

Quel type de sécurité offre la cryptographie? Comment cette sécurité est-elle mesurée? Comment les déploiements de la cryptographie dans le monde industriel ont-ils prouvés leur efficacité? Ce sont des questions importantes. L'article de Diffie et Hellman cité plus haut est remarquable pour la richesse de concepts qu'il introduit. Un de ceux-ci était la promesse que la sécurité des mécanismes cryptographiques pourraient être *prouvée mathématiquement*. La cryptographie a énormément profité de cette notion de preuve, même si dans la pratique, les choses se sont révélées un peu plus compliquées, et l'objectif d'une sécurité prouvable reste toujours quelque peu fugace.¹⁷ Il y a de plus une présomption dans le monde juridique que la preuve mathématique est la démonstration irréfragable d'un fait — c'est ce qui donne tant de conviction aux méthodes cryptographiques.

Dans l'univers cryptographique, la mesure de sécurité dominante concerne la taille de la clé : qui n'a pas entendu de ces savants débats sur les mérites respectifs de la clé de 128, 1024, ou 2048 bits? En mettant l'accent sur la taille de clés, les cryptologues proposent implicitement un modèle de risque, celui d'un adversaire utilisant la force brute de calcul pour briser le système cryptographique. Cette mesure est, selon moi, une des grandes faiblesses de la cryptographie : toute absorbée à ses mathématiques, elle a négligé l'aspect terrain de son domaine, avec comme conséquence que les systèmes cryptographiques ne sont pas brisés par des savantes manipulations mathématiques,

16. Si la définition de certains des blocs fondamentaux de la cryptographie — les *primitives* — font l'objet d'un certain consensus, ce n'est pas le cas pour tout ce qui touche aux interactions et à la composition de ces primitives, que l'on ne sait même pas encore modéliser correctement.

17. Ceci est une discussion qui nous amènerait rapidement dans des eaux presque philosophiques, mais il demeure que la notion même de preuve en cryptographie est loin d'être simple — par exemple, la preuve dans le modèle dit de « l'oracle aléatoire » est considéré par certains comme une forte présomption, plutôt qu'une preuve comme telle.

mais bien par des bidouillages beaucoup moins glorieux.¹⁸

Une autre mesure viendra prochainement s'ajouter à l'évaluation de la vérification des signatures — celle d'une mesure de confiance en la validité d'une clé publique. En effet, l'origine d'une clé publique est toujours affaire d'un processus probabiliste d'évaluation de sa source (« Je fais confiance aux méthodes de Certinomis pour s'assurer de l'identité du détenteur de cette clé »), et la question de la certification des clés devient alors une instance du problème de la prise de décision dans des conditions d'incertitude.¹⁹ Cette incertitude est plus forte lorsqu'il s'agit d'évaluer des certificats provenant de sources étrangères et/ou inconnues. Des chercheurs comme Reiter and Stubblebine suggèrent qu'il faut développer une métrique de confiance, une mesure à appliquer à un certificat, selon le degré de confiance qu'on lui accorde. Même en utilisant des méthodes cryptographiques, l'évaluation de la validité d'une signature serait un processus *probabiliste* et non purement déterministe.

Au-delà de la cryptographie prise isolément, force est de constater que *la sécurité et l'efficacité des technologies d'infrastructures à clés publiques est systématiquement surestimée*.²⁰ Rien ne dit que ces technologies ne rempliront pas leur promesses, mais il faut garder à l'esprit que nous ne disposons d'aucune réalisation qui permette d'évaluer la performance de ces infrastructures sur une échelle substantielle — impliquant plusieurs organisations et plusieurs millions (ou tout au moins, centaines de milliers) d'utilisateurs.²¹ Pour les projets existants, la propension des intervenants (à chaque niveau de la chaîne décisionnelle) à exagérer les chiffres est endémique, car nous fonctionnons dans un marché où il faut acquérir le maximum de capital symbolique, et où être le premier sur les lieux (*first to market*) procure un avantage stratégique déterminant.

Ceci n'est pas nécessairement un état de chose qui doit nous désoler car, après tout, toute technologie mûrit, de même que nos capacités à l'exploiter optimalement. Cependant, les très rares études réalisées sur le terrain tendent à démontrer que si la fiabilité des *algorithmes mathématiques* sur lesquels reposent les procédés de signature électronique semble assurée, ceux-ci ne représentent qu'une faible partie de l'ensemble du processus. L'exemple est typiquement donné de la gestion de clés, une opération hautement délicate et souvent mal comprise : peut importe la robustesse des algorithmes mathématiques utilisés, ceux-ci seront vulnérables dès lors que la gestion des clés — c'est-à-dire les opérations de création, distribution, conservation, destruction — est incorrectement réalisée.²²

Ainsi, dire que l'on dispose aujourd'hui d'*algorithmes* de chiffrement ou de signature fiables

18. Même Whitfield Diffie, grand chantre de la cryptographie moderne, constatait, lors d'un discours, que la cryptographie est « étonnamment difficile à réussir en pratique. »

19. Voir par exemple Reto Kohlas et Ueli Maurer, « Confidence valuation in a Public-Key Infrastructure based on uncertain evidence », in *Proceedings of Public Key Cryptography 00*, Lecture Notes in Computer Science, vol. 1751, pp. 93-112, Janvier 2000.

20. Je ne suis pas le seul à le croire — plusieurs cryptologues réputés ont exprimé cette opinion, par exemple, Ross Anderson, du Laboratoire de recherche sur la sécurité de l'Université de Cambridge, et Bruce Schneir, auteur d'un ouvrage de référence en cryptographie, et plus récemment, de *Secret and Lies*.

21. Un vice-président de la société *RSA Security* confessait récemment que les réalisations d'infrastructures à clés publiques n'ont toujours pas évoluées au-delà de ces vitrines dont on abreuve sans cesse conférenciers et investisseurs.

22. A ce sujet, on consultera avec horreur et bonheur l'article de Ross Anderson, « Why cryptosystems fail », *Communications of the ACM* 37:11 (novembre 1994), pp.32-40.

est une affirmation qu’endosserait une proportion respectable des scientifiques spécialistes de la question ; dire que l’on dispose de réalisations fiables d’infrastructures de sécurité — c’est-à-dire permettant la génération, certification, distribution, mise à jour automatique et révocation de clés de signature et/ou de chiffrement, les services d’annuaire, les mises à jour d’annuaire, l’archivage, l’audit, etc. — est une affirmation qui relève de la seule science du ... marketing.

3.1.6 Évaluation

Clairement, la signature cryptographique offre des avantages importants pour la sécurisation des réseaux. En particulier,

- 1° Elle garantit fortement l’intégrité des données qui transitent à travers les réseaux ;
- 2° Elle garantit un fort lien entre **clé** et **document** ;
- 3° Elle est susceptible d’être automatiquement générée et vérifiée, sans intervention humaine et, de ce fait, idéale pour sécuriser les communications entres machines ;

Cependant, dans le cadre de notre réflexion sur l’acte authentique, le modèle cryptographique de la signature pose des problèmes importants. En particulier,

- 1° L’héritage militaire de la cryptographie a imposé un modèle où l’environnement est nécessairement et maximalelement hostile. Ce modèle implicite a pour effet de dévaloriser le rôle des institutions dans la sécurité juridique ;
- 2° La sécurité de la cryptographie est toujours évaluée en fonction de taille de clés. Cette mesure est peu pertinente dans un cadre opérationnel ;
- 3° Le modèle cryptologique est ergonomiquement pauvre, ne tirant aucunement partie de l’héritage culturel de la signature manuscrite et ne préservant pas les solennités de l’acte. En particulier, la signature cryptologique ne dispose d’aucune représentation visuelle ;
- 4° Une signature cryptographique n’identifie jamais qu’une clé privée, et non un individu ; Ainsi, le « contrôle exclusif du procédé de signature » requis par la Directive n’est assuré, en bout de compte, que par le code personnel à quatre chiffres qui contrôle l’accès à la clé entreposé sur la carte à puce ;
- 5° La signature cryptographique nécessite une infrastructure extrêmement lourde, les ICPs ;
- 6° La signature cryptographique exige que toute partie à un acte soit enregistrée au sein du système pour pouvoir signer ;
- 7° Le modèle cryptographique impose que les signatures soient toujours vérifiées.

Il faut donc considérer attentivement les nouveaux risques associés à l’utilisation de la signature cryptologique. Avant tout, il faut se départir de l’idée que la signature cryptologique représente une « grande et belle » signature électronique, opposé à d’autres formes de signature qui seraient, elles, des « ersatzs ».

3.2 La signature biométrique

La signature biométrique se fonde sur une toute autre approche que celle de la cryptologie, tant dans sa définition de la signature, que dans sa mesure de sécurité et son mode de preuve, que dans

l'appareillage requis. La biométrie se rapporte à la mesure de *caractéristiques physiques* uniques à l'individu — la plus connue étant évidemment l'empreinte digitale. De nombreuses mesures ont été développées, chacune pourvues de caractéristiques différentes : parole, rétine, iris, géométrie de la main, et même l'odeur!²³ Commençons par dissiper un certain nombre de malentendus à propos de la signature biométrique.

D'emblée, notons qu'une mesure biométrique prise par elle-même n'est pas une signature — on entend souvent que « bientôt nous pourrons signer avec notre œil! (ou son pouce) ». Alors que la signature est couramment utilisée comme marque de manifestation de volonté (j'approuve, je m'engage, j'ai lu, j'ai noté, etc) sans être une marque d'identité — comme dans le contrat notarié — l'inverse n'est pas vrai, c'est-à-dire qu'une marque d'identité seule ne suffit jamais à manifester une intention. Or les mesures biométriques seules, *sans contextualisation* ne sont pas la manifestation d'une intention. En théorie, rien n'empêche que soit développée une technologie — avec *le contexte culturel dont elle dépend* — qui permette de décréter que l'exposition de sa rétine à un lecteur devienne une forme de signature, si ce n'est que depuis un millénaire, la manifestation du consentement contractuel est accompagné de *gestes*, d'une expression corporelle quelconque.

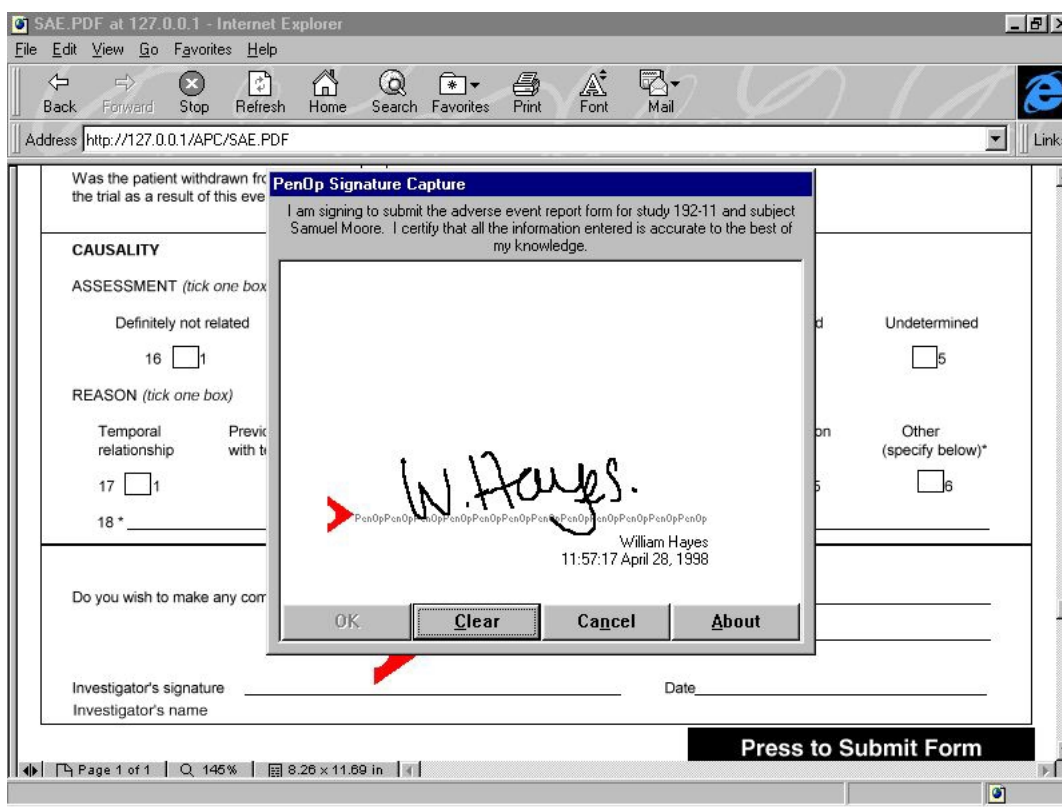


FIG. 3.1 – Signature biométrique à l'aide du produit PenOp et d'une tablette graphique.

23. Chacune de ces techniques présente des avantages et inconvénients — pour un survol, voir le rapport de Dirk SCHEUERMANN, *Usability of Biometrics in Relation to Electronic Signatures*, EU Study 502533/8, GMD Forschungszentrum Informationstechnik GmbH.

Il faut de plus distinguer entre l'identification biométrique et la signature biométrique, car elles font appel à des opérations techniques différentes : dans la signature, on ne cherche pas à *déterminer* l'identité, mais à la *corroborer*. C'est-à-dire qu'un individu déclare son identité, produit sa signature, et on vérifie si cette signature correspond bien à celle associée à l'individu. Le même principe est à l'oeuvre pour la signature cryptographique : lorsqu'on envoie un message signé, le destinataire ne détermine pas l'identité du signataire à partir de la seule signature, en fait, il en serait bien incapable ! Plutôt, il récupère le certificat à clé publique du signataire, et *corrobore* l'identité supputée par le processus de vérification. Tout comme, dans un acte, on obtient l'identité du signataire par le fait que son nom est tapé en toutes lettres au-dessus de signature.²⁴

Finalement, les méthodes biométriques sont reliées au problème de la signature électronique de deux façons : (1) dans le contexte de la signature cryptographique, l'ensemble des méthodes biométriques peuvent être utilisées pour assurer le lien entre un individu et sa clé privée, remplaçant par exemple le code personnel à quatre chiffres ; (2) une technologie biométrique particulière, *l'analyse dynamique de la signature manuscrite*, est en soi une méthode de signature électronique. Le premier cas, qui correspond à une technologie de *contrôle d'accès*, ne sera pas examiné dans ce rapport — encore une fois, le lecteur est prié de se référer à l'excellent rapport de Dirk SCHEUERMANN pour une discussion assez complète des avantages et des inconvénients de chacune des méthodes biométriques, dans une perspective d'utilisation pour le contrôle d'accès à une clé privée.

3.2.1 Mécanisme de la signature biométrique

La particularité de l'acte de signer est que chacune des signatures d'un individu est différente et pourtant, l'identifie uniquement. Les technologies de signature biométrique tire parti de ces caractéristiques pour produire un système qui permet tout à la fois d'améliorer les techniques établies de vérification de signature et de conserver une forme de manifestation du consentement — écrire son nom de sa main — qui jouit d'une acceptation presque totale au sein des cultures occidentales. Un système de signature biométrique consiste en plusieurs opérations : l'enregistrement, la signature et l'identification et l'intégrité.

Enregistrement : Pour signer, l'utilisateur doit disposer d'une tablette graphique et d'un stylet approprié, et le système doit disposer d'un patron. Selon le produit spécifique de signature utilisé, le logiciel effectue différentes mesures lors de la signature : angles x , y et z , pression, vitesse. Ces mesures sont ensuite comparées à un *patron*, c'est-à-dire une sorte de moyenne de la signature d'un individu. Ce patron peut être obtenu soit (1) *avant* la signature, soit (2) *après*.

— (1) Lorsqu'un individu compte utiliser régulièrement le système, sa signature est enregistrée. Cette opération consiste en l'obtention d'un certain nombre d'exemplaires de sa signature — typiquement, de 3 à 10 — de façon à établir un *patron* de sa signature, c'est-à-dire une moyenne des différentes mesures effectuées à partir de la signature : coordonnées x , y et z , pression, vitesse. Le patron est ensuite conservée au sein d'une base de données, pour être utilisé, en temps réel, lors

24. Cette distinction est importante techniquement : un système biométrique d'identification doit comparer la mesure biométrique avec l'ensemble des patrons des candidats — par exemple, dans le cas des empreintes digitales recueillies sur la scène d'un crime, on compare la mesure avec l'ensemble du fichier d'empreintes de la police, une procédure très laborieuse, évidemment. La corroboration est beaucoup plus efficace, puisqu'elle n'a qu'une seule comparaison à effectuer, avec le patron de l'individu concerné.

de la vérification d'une signature.

— (2) Dans ce cas, on obtient les exemplaires requis de la signature de l'individu *après* une contestation. Cette type de situation se présente en fait dans la plupart des cas où il n'existe pas de lien préalable entre la personne et le système. La signature n'est pas vérifiée lors de la signature du document, mais les mesures sont simplement conservées avec le document. Ce n'est que lorsque la validité de la signature est contestée que l'on se procure les exemplaires de signature nécessaire à la construction du patron. Ceci reproduit ce qui se passe dans la réalité : la grande majorité des signatures n'est jamais vérifiée, c'est-à-dire que *la signature est uniquement vérifiée en cas de litige* plutôt que systématiquement.

Signature, identification : La signature proprement dite est effectuée à l'aide d'un logiciel intégré à l'application produisant le document (Microsoft Word, ou Adobe Acrobat par exemple). Les mesures effectuées sur la signature sont alors comparées avec le patron. Quatre résultats sont possibles : (a) la signature est jugée conforme au patron, et justement acceptée ; (b) la signature est jugée non-conforme, et justement refusée ; (c) la signature est jugée conforme, et injustement acceptée ; (d) la signature est jugée non-conforme, et injustement refusée. Les algorithmes de vérification de signature doivent tenter de maximiser (a) et (b) et de réduire au maximum (c) et (d). On appelle (c) le taux d'*acceptation erronée* et (d) le taux de *rejet erroné*.

Lien et intégrité : Tout comme la plupart des mécanismes de signature cryptographique, la signature biométrique établit le lien entre la donnée de signature (clé privée, mesure biométrique) et le document par l'utilisation d'une fonction de condensation (ou fonction de hachage) — ceci permet de lier la signature à un message donné, et d'assurer que, de façon pratique, toute altération du message entraîne l'échec de la vérification de la signature.

3.2.2 Quantification, normalisation

Les algorithmes qui effectuent la comparaison entre les mesures prises lors de l'exécution d'une signature et le patron conservé au sein du système retournent une *mesure statistique de confiance*. Puisque le procédé biométrique est fondé sur la propriété que chaque exécution d'une signature est différente de toutes les autres, tout en étant unique à chaque individu, la mesure de comparaison ne peut être exacte. En fait, une correspondance exacte entre le patron et une signature donnée serait à coup sûr le signe d'une tentative de fraude. La signature biométrique est donc, en son essence, probabiliste.²⁵ Ainsi, lors du processus de vérification de la signature, les produits de signature biométrique offrent une mesure de validité de la vérification, laissant à l'utilisateur l'option de déterminer si le contexte d'utilisation impose une vérification plus ou moins certaine.²⁶

Ceci dit, à quoi correspond cette mesure de validité ? Malheureusement, la situation n'est pas claire. Comme il n'existe aucune norme — *de jure* ou *de facto* — régissant les algorithmes de vérification de signature biométrique, il est difficile de les comparer entre eux. Pire encore, il n'existe

25. Dans le modèle cryptographique, l'identification de la clé qui a signé le document est déterministe, mais c'est le lien entre individu et clé qui introduit une dimension probabiliste au processus.

26. Ceci reproduit assez bien ce qui se produit lorsque l'on signe un chèque important et que l'on doit s'appliquer à « bien » signer, c'est-à-dire à apposer une signature qui ressemble à celle déposée à la banque lors de l'ouverture de notre compte chèques.

pas non plus de normalisation des mesures que l'on pourrait appliquer pour déterminer les mérites respectifs des algorithmes. La seule mesure communément utilisée est le FA/FR discuté plus haut, mais cette mesure est largement illusoire dans le contexte de la sécurité, puisque il faudrait également déterminer ce que constitue une attaque, quelles sont les moyens à la disposition d'un faussaire, son accès à des exemplaires de signature manuscrite, etc.²⁷

L'absence de quantification et de normalisation est-elle nécessairement un handicap? Il faut réfléchir attentivement à cette question. Indépendamment de cette question, la communauté biométrique a encore beaucoup à faire avant de pouvoir en arriver à un niveau de normalisation comparable à celui de la communauté cryptographique.

3.2.3 Évaluation

La signature biométrique présente des caractéristiques intéressantes pour un grand nombre de contextes :

- 1° Elle est culturellement ergonomique, en ce sens qu'elle présente une variation minimale de la manifestation du consentement telle qu'on la réalise dans les sociétés occidentales, par la signature de son nom ;
- 2° Du point de vue de l'expertise judiciaire, elle permet de conserver une forme éprouvée de mesure, l'analyse statique d'une signature, et d'y ajouter la mesure dynamique, c'est-à-dire les paramètres de vitesse et de pression. Son expertise s'inscrit au sein d'une tradition d'expertise des signatures aux modalités déjà bien connues ;
- 3° Elle apporte une certaine sécurité sans nécessairement imposer la vérification automatique de la signature ;
- 4° Elle produit une représentation visuelle de la signature ;
- 5° Contrairement à un code d'accès, on n'oublie pas sa signature.

Par contre,

- 1° Elles peuvent être sensibles aux modifications de l'état de la personne, i.e, par exemple, ingestion d'alcool, maladie ou âge provoquant un tremblement de la main (conditions qui peuvent tout aussi bien entraîner l'oubli du code personnel) ;
- 2° Une fois qu'il est compromis, le patron biométrique ne peut se renouveler! C'est-à-dire que les caractéristiques dynamiques qui identifient uniquement la signature d'un individu ne sont pas observables directement, mais ils ne sont pas non plus renouvelables à loisir, comme l'est un mot de passe, ou une clé cryptographique.

3.3 la signature-tatouage

A celui qui désire protéger un secret, deux approches sont possibles: le chiffrement et la dissimulation. Le chiffrement cryptographique permet de rendre un secret inaccessible, même si le

27. Voir à ce sujet Nick Mettyear, « Error Rates in Biometric User Authentication », Mémoire PenOp.

texte crypté est en soi accessible ; les techniques de *dissimulation d'information*²⁸ tentent, elles, de rendre le secret **invisible**.²⁹ Cette approche peut s'imposer dans certaines situations, où l'utilisation de la cryptographie est *en soi* un aveu que l'on dispose d'un secret.

Historiquement, la dissimulation d'information a connu de nombreuses formes et applications, mais, récemment, cette science a connu un renouveau, alors qu'est apparu l'intérêt de son application à la protection des œuvres numériques. Dans le contexte d'œuvres infiniment copiables, sans dégradation de qualité, à la distribution instantanée et sans frontières géographiques, les technologies de *tatouage*³⁰ offrent la promesse de pouvoir apposer des « marques » indélébiles aux œuvres numériques, marques qui permettraient de retracer soit les ayant-droits, soit les propriétaires légitimes des ces œuvres. Le tatouage est appliqué directement à l'oeuvre et réside en son sein — c'est-à-dire que le signal encodant le tatouage est intimement mêlé au signal de l'oeuvre elle-même, rendant sa découverte ou son extraction difficile pour ceux ne disposant pas des paramètres ayant permis le tatouage original.

Les tatouages se trouvent en deux saveurs : *fragiles* ou *robustes*. Dans le premier cas, toute manipulation de l'image entraîne la disparition du tatouage — en d'autres termes, la présence du tatouage témoigne de l'intégrité du document. Dans le second cas, éliminer le tatouage exige d'endommager irrémédiablement l'oeuvre au point de la rendre inutilisable, ou à tout le moins de réduire significativement sa qualité. On peut diviser les tatouages robustes en *tatouages visibles* et *tatouages invisibles*.

Les *tatouages visibles* sont un équivalent des marques en filigranes que l'on retrouve sur certains documents sécurisés, le papier à en-tête, ou sur les papiers spéciaux (papier-monnaie, passeports). Ces marques couvrent une grande surface du papier, mais demeurent semi-transparentes, de façon à ne pas affecter la lisibilité du document. Dans le contexte d'une image numérique, elles permettent d'identifier le producteur d'une oeuvre. La figure 3.2 montre un tatouage visible appliquée à une image numérique, obtenue d'un livre appartenant de la Bibliothèque du Vatican. Le tatouage identifie l'origine du document, tout en n'entravant pas sa lisibilité. Toute tentative d'éliminer le tatouage entraîne une dégradation de la qualité de l'image, idéalement.

28. *Information hiding*.

29. Note terminologique : le champ scientifique de la dissimulation d'information étant encore très jeune, une terminologie globalement acceptée émerge à peine. Une excellente tentative de classification des nombreux concepts du domaine est celle de Fabien A. P. PETITCOLAS, Ross J. ANDERSON et Markus G. KUHN, « Information Hiding — A Survey », *Proceedings of the IEEE*, 87(7): 1062–1078, juillet 1999.

29. Le tatouage des œuvres, numériques ou non, permet de réaliser des objectifs surprenants en terme d'authentification et d'identification. Par exemple, Margaret Thatcher, excédée par les nombreuses fuites de documents confidentiels, utilisa un système qui permit d'identifier la source des fuites ; le système consistait à encoder un identifiant dans chaque copie initiale du document par le décalage d'un $1/300^e$ de pouce des lignes du texte — un décalage vers le haut ou vers le bas permettant d'encoder un chiffre binairement. L'intérêt d'un tel système est qu'il est invisible et peut résister à plusieurs copies successives de même qu'à la télécopie. La copie « coulée » aux journaux pouvait donc être retracée jusqu'à l'auteur original de l'indiscrétion. Évidemment, la détection peut être évitée par simple retranscription du document, mais encore fallait-il être conscient du marquage du document. Voir J. BRASSIL, S. LOW, N. MAXEMCHUK, et L. O'GARMAN, « Electronic marking and identification techniques to discourage document copying », *Infocomm*, pp. 1278–1287, IEEE, juin 1994.

30. *Watermarking*.

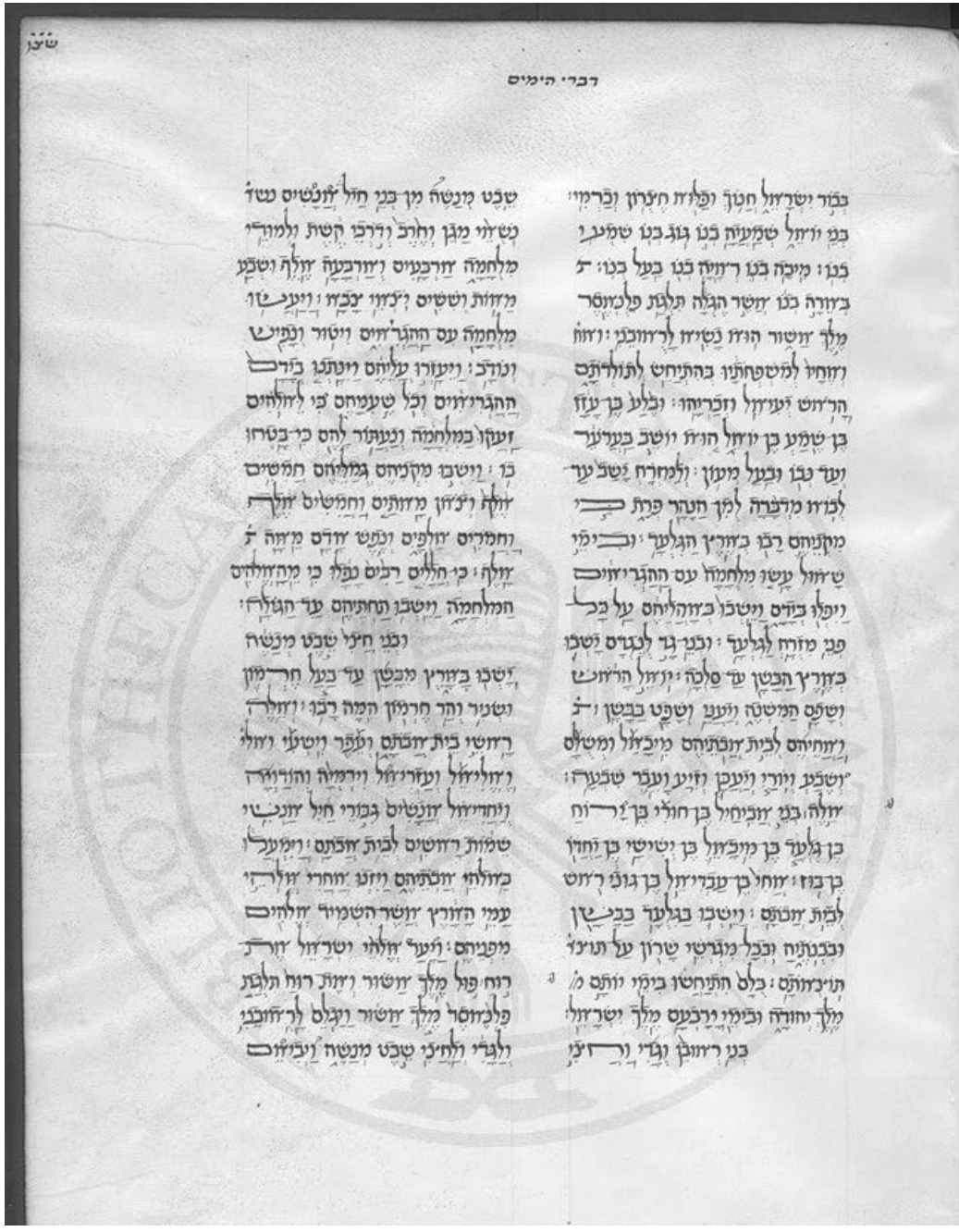


FIG. 3.2 – Un tatouage visible d'un document numérisé de la Bibliothèque du Vatican. Le tatouage n'est pas sur le papier, mais bien au sein du fichier informatique.

Les *tatouages invisibles* ne sont pas perceptibles à l'œil. Ils consistent simplement en un identificateur quelconque inséré au sein d'une image. Le procédé d'insertion assure qu'on ne peut supprimer cet identificateur sans irrémédiablement endommager l'image. Pour s'assurer de l'intégrité d'une image, on dispose d'un procédé qui vérifie la présence de l'indentifier au sein de l'image. Remarquablement, cet indentifier survit au passage du numérique à l'analogue, i.e., le

papier. En d'autres termes, un document contenant une tatouage invisible peut être imprimé (par laser, jet d'encre, etc) et il sera toujours possible à l'algorithme de validation de retrouver l'identificateur, après numérisation du document papier. Ceci est un avantage notable sur les autres technologies de signature qui perdent toute capacité de vérification dès l'instant où le document est imprimé sur papier.

3.3.1 Mécanisme de la signature-tatouage

Comment le marquage des œuvres fonctionne-t-il au niveau de la signature? En utilisant une combinaison de technologies qui, ensemble, permettent de réaliser les trois fonctions de la signature, telle que décrite par la loi du 13 mars : identification, intégrité, lien avec le document.

Par exemple, la technologie VeriData, développée par la société Signum Technologies,³¹ permet de signer un document de la façon suivante :

- 1° Un condensé de l'image est calculé, à l'aide d'une fonction de hachage cryptographique ; ce condensé est chiffré à l'aide d'un algorithme de chiffrement symétrique ;
- 2° L'image est divisé en un certain nombre de rectangles ;
- 3° Chaque signataire dispose d'une clé — un simple code personnel ; cette clé est utilisée pour déterminer un certain nombre de locations au sein de chaque rectangle ;
- 4° La valeur du condensé chiffré est insérée au sein de l'image en ces locations, en modifiant la valeur de luminosité ou la valeur chromatique des pixels ;
- 5° Pour valider le document, on doit disposer de l'image originale, et de la clé, de façon à vérifier la valeur des condensés aux locations déterminées par la clé. Si la vérification échoue, on sait que l'image a été modifiée, mais on sait en plus dans quelle région de l'image.

La mesure d'intégrité est évidente — notez qu'elle est encore plus fine que celle fournie par la signature cryptographique ou biométrique, puisque qu'elle permet, le cas échéant, de préciser quelle région de l'image a subi des modifications. Le lien de l'identificateur avec le document est également évident puisque cet identificateur détermine la location des condensés insérés au sein de l'image. Selon le mécanisme de gestion de la clé de signature, celle-ci peut être considérée comme identifiant uniquement le signataire.

3.3.2 Quantification, normalisation

La recherche est encore très jeune en ce domaine, et on peut s'attendre à voir émerger rapidement des solutions de plus en plus sécuritaires et adaptées à des contextes de plus en plus variés. La très forte demande pour de tels produits dans le contexte de la protection de la propriété intellectuelle va pousser vers une très grande activité scientifique et d'innovation technologique. Cependant, le domaine, tout comme celui de la signature biométrique, ne s'est pas encore entendu sur des algorithmes communs, et les sociétés fondent leur avantage commercial sur les différences de performance entre les différents procédés proposés. Au niveau de la mesure et de la quantification, tout reste à faire.³²

31. <http://www.signumtech.com>.

32. Fabien Petitcolas a commencé à débroussailler le terrain de façon remarquable — voir le site Web qu'il consacre à la stéganographie : <http://www.cl.cam.ac.uk/~fap>.

3.3.3 Sécurité

La question de la sécurité des procédés de tatouage numérique est épineuse et fait l'objet de nombreux débats au sein de la communauté des experts en sécurité. Deux conceptions de la sécurité s'affrontent : une veut que les procédés de sécurité soient examinés au sein de la communauté d'experts, cette méthode étant la seule susceptible de fonder notre confiance ; l'autre prétend que rien ne sert de fournir aux fraudeurs plus d'informations qu'il n'est nécessaire. La Banque de France ne publie pas sur son site Web la méthode de fabrication du papier-monnaie après tout, tout comme Canal+ ne publie pas les plans de son décodeur. La première conception est celle des chercheurs en cryptographie, alors que la seconde semble, pour l'instant, caractériser les procédés de tatouage de oeuvres. Il est donc difficile de discuter de la sécurité de procédés de tatouage — tout dépendra de l'évolution du marché et de la réglementation.³³

Les procédés de tatouage seront sans doute utiles pour s'assurer qu'un utilisateur lambda ne puisse frauder, mais ils auront peu d'effet envers un attaquant déterminé. Ces procédés doivent donc être intégrés au sein d'autres mécanismes de sécurisation — tout comme les cartes bancaires ou les passeports utilisent toute une panoplie de mesures (hologrammes, polymères spéciaux, filigranes, etc.) pour freiner les faussaires.

3.3.4 Évaluation

La signature-tatouage présente donc des caractéristiques qui la différencient des autres technologies de signature électronique :

- 1° Elle est entièrement contenue au sein de l'image — pas de méta-données à gérer, elle est littéralement liée au document signé ;
- 2° Sa notion d'intégrité est plus fine, car elle permet de localiser l'endroit où le changement a eu lieu dans le document, plutôt que de fournir une simple réponse binaire ;
- 3° Elle survit au passage du numérique à l'analogique et vice-versa.

D'autre part,

- 1° Les technologies ne sont pas encore bien testées et stabilisées.

La signature-tatouage pourrait donc être mise en oeuvre dans le contexte d'ajout de niveaux de sécurisation supplémentaires au sein des images — possiblement par les pilotes d'impression et de numérisation, de façon à insérer directement au sein de l'image des informations de traçabilité. Ces marques, visibles ou invisibles, en « filigrane » ne seraient pas nécessairement vérifiées, mais fourniraient des sécurités supplémentaires, en cas de contestation de l'authenticité du document. Les indications de traçabilité que fourniraient ces marques sur la provenance ou la date d'un document, bien qu'elles soient pas nécessairement assimilables à une signature, s'ajouteraient utilement au « faisceau de preuves » à partir duquel l'intégrité d'un document est susceptible d'être établie.

33. En effet, aux Etats-Unis, le *Digital Millennium Copyright Act* interdit purement et simplement de briser les mécanismes de protection des oeuvres, indépendamment de leur sécurité : section 1201(a)(1) débute par : « Aucune personne ne peut contourner un mécanisme technologique qui a pour effet de contrôler l'accès à une oeuvre protégée par ce texte. » (*No person shall circumvent a technological measure that effectively controls access to a work protected under this title.*) Plutôt que d'investir dans la recherche, il est peut-être plus simple de criminaliser la production et la distribution de mécanismes de fraude, mais il est encore trop tôt pour voir si cette approche sera efficace.

3.4 La signature numérisée

La signature numérisée consiste simplement en la capture, au sein d'un fichier informatique, de l'*image* de la signature manuscrite d'un individu. L'image informatique résultante peut ensuite être ajoutée, par différents procédés, à la suite ou au sein d'un document électronique.

Avant de pouvoir discuter des mérites de cette technologie, il me faut tout d'abord prendre un peu de recul et justifier son inclusion dans ce rapport, puisque sa simple évocation mène inévitablement à une violente levée de boucliers : « Mais voyons ! Un tel procédé ne peut être qualifié de signature électronique ! La signature apposée sur un tel document peut être contrefaite par une simple opération de couper-coller ! » Une telle argumentation procède d'une analyse erronée du cycle de vie d'un document, analyse qui postule que les documents électroniques sont créés, distribués et lus uniquement sous forme électronique. Hors, ceci ne correspond qu'à une infime partie des usages observés dans la pratique : en fait, l'essor de l'outil informatique a stimulé plus que jamais l'usage du papier et des technologies qui assurent la médiation entre le support papier et l'électronique — télécopieurs, imprimantes, scanners.³⁴ De plus, il y a intensification des connexions entre ces outils : les photocopieurs font aussi emploi d'imprimantes réseau, on peut envoyer des courriers électroniques à des télécopieurs connectés à l'Internet et, inversement, des télécopies à un ordinateur. Il n'y a donc pas une « logique du document électronique », fruit de la marche victorieuse et inlassable du progrès technologique, qui supplanterait une « logique du document papier », synonyme d'une civilisation désuète désormais vouée à l'extinction. Il y a plutôt une interaction complexe et synergétique entre deux supports, interaction qui les dynamise mutuellement.³⁵

Concevoir l'informatisation des opérations bureaucratiques en tenant compte de cette réalité permet d'éviter de s'enliser dans le dogme du « tout électronique » et de concevoir des solutions technologiques qui intègrent et tirent parti de cette nature hybride du document.

3.4.1 Évaluation

Le système SAGA développé pour le Service central de l'état civil (SCEC) à Nantes nous permettra d'évaluer la signature numérisée dans son contexte global d'utilisation. Le système développé pour le SCEC est remarquable tant pour la simplicité de conception que pour son adéquation aux besoins exprimés.³⁶ Le système visait à permettre aux officiers d'état civil du SCEC de délivrer le plus rapidement possible des copies conformes d'actes d'état civil, une activité qui occupait une portion de plus en plus importante de leur temps.³⁷ Ces copies sont remises soit directement au particulier, soit à des notaires ou autres institutions requérantes.

Le système comporte trois éléments distincts : tout d'abord, les 8 millions d'actes numérisés à

34. Voir par exemple « What paperless office? Fax usage is up », *Managing Office Technology* 42:1(39).

35. Voir à ce sujet Ziming LIU et David G. STORK « Is Paperless Really More? Rethinking the Role of Paper in the Digital Age » *Communications of the ACM* 43:11(94-97).

36. Si vous ne croyez pas que ce soit remarquable, c'est que vous n'avez pas encore assez fréquenté d'informaticiens. . .

37. Voir le compte-rendu de Mme BANAT-BERGER pour une description générale du système et du contexte institutionnel — <http://www.gip-recherche-justice.fr/preuve/etatcivilnantais.htm>.

partir des registres papier³⁸ ; ensuite, le système qui permet à l'officier d'état civil d'apposer un « pavé » contenant le sceau de l'État et sa signature au sein de l'acte numérique (voir la figure 3.3) ; finalement, le papier sécurisé sur lequel la copie conforme signée est imprimée, papier pourvu de caractéristiques spéciales qui protège son intégrité et en empêche la reproduction.³⁹

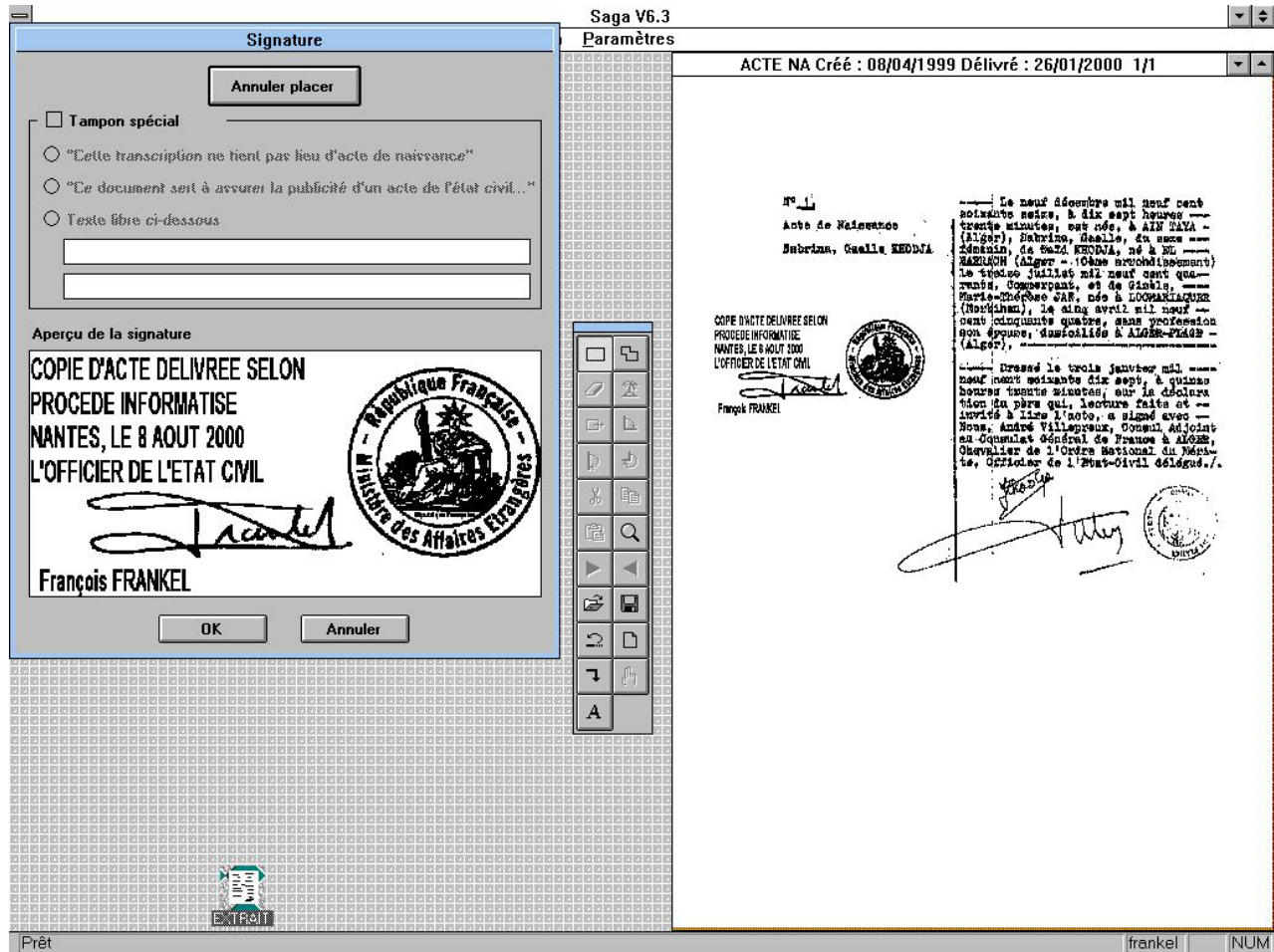


FIG. 3.3 – Application de la signature de l'officier d'état civil à un acte numérisé — logiciel « SAGA » du Service central d'état civil de Nantes. A gauche, la boîte de dialogue contenant le sceau, la signature, la mention « Copie d'acte ... », la date, etc. ; à droite, le document numérisé avec la signature de l'officier, prêt à être imprimé.

Au sein même du SCEC, la sécurité est assurée par un ensemble de méthodes et techniques : les officiers d'état civil ne sont pas n'importe quel groupe d'utilisateurs, et la pénalisation sévère du faux en écriture publique assure, plus que toute mesure technologique ne saurait le faire, qu'ils

38. La capture des actes sous forme de fichier image est la seule concevable compte tenu de la nature des documents — à moins de considérer la retranscription manuelle des 8 millions d'actes ...

39. Pour un survol des technologies de sécurisation matérielle des documents, voir Rudolf VAN RENESSE, ed. *Optical Document Security*, Artech House Publishing, 1998.

ne soient pas considérés comme des fraudeurs potentiels au sein du système.⁴⁰ Des mécanismes de journalisation automatique des procédures informatiques assurent leur traçabilité. L'accès aux locaux contenant les documents numérisés et au papier sécurisé est contrôlé par les procédés traditionnels de clés et serrures, et l'accès aux postes de travail des officiers d'état civil est quant à lui contrôlé par mot (ou phrase) de passe. De plus, l'utilisation de la signature numérisée de l'officier est strictement guidée par le système : l'image numérisée de la signature ne réside pas sur le poste de travail de l'officier mais bien sur un serveur central, n'étant transférée qu'au moment de l'identification de l'officier à son poste de travail ; la signature n'est utilisable que dans les modalités définies par le système (comme celles visibles à la figure 3.3) et ne peut être extraite pour une utilisation non-conforme.

La solution développée par le SCEC remplit tout à fait les conditions pour la signature énoncées par la loi du 13 mars (article 1316-4, alinéa 2) : elle fournit une identification fiable de l'officier d'état civil ayant apposé sa signature sur l'acte, à la fois par la vérification traditionnelle des signatures et par les mécanismes de journalisation qui permettent de retracer les actions sur les actes ; elle garantit le lien avec l'acte auquel elle s'attache, puisque l'acte signé n'est accessible que sur un support papier sécurisé qui assure l'origine et l'intégrité de l'acte.

La solution développée par le SCEC est utilisable *dès à présent* et s'intègre directement dans l'univers bureaucratique *tel qu'il existe aujourd'hui*. En préservant les caractéristiques visuelles des documents d'état civil (sceau, signature), elle s'intègre sans effort aux schémas cognitifs des différents intervenants. Cette solution tire donc son efficacité d'une analyse adéquate du cycle de vie du document et de son réseau de distribution — et non de l'application d'une solution technologique déterminée préalablement à toute évaluation des problèmes à résoudre. Si l'on considère qu'un tel système intégrera éventuellement la distribution des actes par voie électronique, il est clair qu'il faudra alors utiliser des technologies supplémentaires pour assurer l'intégrité des documents en transit — je suggère à cet égard des voies de réflexion à la conclusion de ce chapitre.

En résumé, lorsque l'on considère *l'ensemble global* des procédures de distribution et de sécurisation des documents, la signature numérisée est un élément important dans la boîte à outils des institutions qui désirent informatiser leur processus de production et de délivrance de documents administratifs.

3.5 Conclusion

Après la description de ces quatre modèles très différents de réalisation de la signature électronique, on remarque que chacune de ces technologies traite le mieux **un aspect précis** de la signature manuscrite :

- 1° La cryptographie établit le mieux **le lien entre signature et document** dans le contexte de la transmission à distance. Par contre, elle ne dispose d'aucune représentation visuelle et n'établit pas de lien direct avec la personne (c'est le rôle de la certification). Elle est idéalement adaptée à la signature automatique par des machines ;

40. Considérez qu'un système de paiement comme celui de la carte bleue doit être protégé des fraudes commises par les consommateurs, mais aussi les commerçants, les employés de banque, etc.

- 2° La biométrie permet de s'assurer de **la présence physique du signataire** lors de la signature et, des quatre technologies, reproduit le mieux la dimension rituelle du consentement ;
- 3° Le tatouage visible et la signature numérisée permettent de traiter **les marques visuelles d'authenticité** comme le sceau et le filigrane, reproduisant, en quelque sorte, l'incrustation de l'encre au sein du papier ;
- 4° Le tatouage invisible permet de réaliser **la traçabilité des documents électroniques**, en incluant au sein de l'image des informations comme la date, le poste de production et le numéro de série.

Il est tout à fait plausible d'imaginer que des solutions technologiques évolueront qui harmoniseront ces approches en des configurations variables, selon la nature des documents considérés. La préservation des qualités visuelles des documents jouera sûrement un rôle important pour les actes qui ont traditionnellement fait usage de la riche palette des signes de l'authenticité — sceau, tampons, filigranes, signature manuscrite, etc. Un rapprochement s'effectuera naturellement entre l'exigence de la présence physique de l'officier public et l'utilisation de la biométrie. La cryptographie s'imposera dès qu'il s'agira d'assurer l'intégrité d'un document transitant électroniquement au sein d'un réseau ouvert, sans qu'on doive nécessairement conceptualiser cette « signature » cryptographique comme celle apposée sur le document.

Cependant, contrairement au dogmes qui circulent, **l'utilisation de la cryptographie à des points précis du système n'impose pas que l'on adopte dans son ensemble la (Sainte) Trinité cryptographie-PKI-carte à puce**. Ce modèle de sécurité impose en effet une harmonisation excessive des pratiques bureaucratiques, ne tenant pas compte des différentes cultures organisationnelles des institutions. Il est plus efficace, moins coûteux et, en bout de ligne, globalement plus sécuritaire de laisser chaque institution déterminer la façon dont elle désire sécuriser la circulation des documents à l'interne. En effet, considérons le modèle de circulation des documents présenté à la figure 3.4 : on y voit deux institutions (origine et arrivée) qui échangent des documents entre elles et avec un individu au sein d'un réseau *ouvert* (c'est-à-dire ouvert à des individus susceptibles de commettre des fraudes). Chaque institution dispose de son propre réseau *fermé* (représenté par le cercle intérieur ombré) — ce qui peut être obtenu par différents moyens techniques (mur coupe-feu, contrôle d'accès, etc.) Pour échanger des documents avec des individus, les institutions utilisent un papier sécurisé assurant un minimum de protection contre la contrefaçon et la reproduction. Pour échanger des documents entre elles, les institutions utilisent une *passerelle* (représentée par le cercle extérieur) : c'est cette passerelle qui assure la traduction entre les technologies de sécurisation utilisées à l'interne et les technologies cryptographiques utilisées pour la communication au sein du réseau ouvert.

Si l'institution d'origine dispose d'une politique de sécurité interne dont la composante principale est l'application de la peine capitale à toute personne tentant de falsifier un document, ceci ne devrait pas avoir d'impact sur l'institution d'arrivée qui peut, quant à elle, décider d'investir dans un système qui soit en accord avec sa propre culture institutionnelle, la nature de documents, les relations avec les partenaires, etc. — par exemple, s'en remettre entièrement à la bonne foi de ses membres.⁴¹ Plutôt que de tenter d'imposer le même modèle de confiance à chacun des éléments

41. La Loi type de la CNUDCI sur le commerce électronique suggérait à l'alinéa 58 une telle considération de l'ensemble des facteurs de sécurité : « Pour déterminer si la méthode utilisée en vertu du paragraphe 1 est appropriée, les facteurs juridiques, techniques et commerciaux à prendre en considération sont les suivants : 1) le degré de perfectionnement du matériel utilisé par chacune des parties ; 2) la nature de leur activité commerciale ; 3) la fréquence avec laquelle elles effectuent entre elles

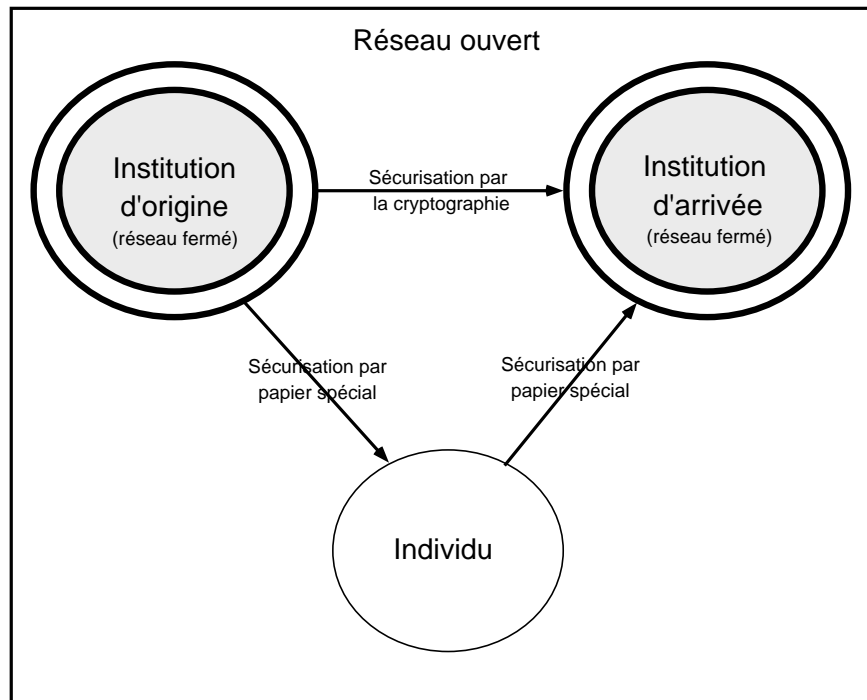


FIG. 3.4 – Circulation d'un document entre institutions et individus

du système, il est beaucoup plus facile et bureaucratiquement écologique d'établir des passerelles qui assurent la traduction entre les différents régimes de confiance et de sécurité propres à chaque institution.

Un telle passerelle pourrait, par exemple, prendre la forme d'une PKI où *une seule paire de clés privée/publique est attribuée à chaque institution*.⁴² Les documents qui doivent circuler à l'extérieur de l'institution sur un réseau ouvert comme l'Internet peuvent être sécurisés par l'utilisation de ces clés (signature et/ou chiffrement), chaque institution étant alors en mesure de vérifier que les documents ont transités au sein du réseau sans subir de modifications. Un tel système réduit de façon considérable les inconvénients des technologies de PKI — comme le foisonnement des clés et des certificats — mais présente évidemment peu d'attrait pour les sociétés de certification puisqu'il va à l'encontre de leur intérêt commercial, qui consiste à vendre (et renouveler) le plus grand

des opérations commerciales; 4) la nature et l'ampleur de l'opération; 5) le statut et la fonction de la signature dans un régime législatif et réglementaire donné; 6) la capacité des systèmes de communication; 7) les procédures d'authentification proposées par les opérateurs des systèmes de communication; 8) la série de procédures d'authentification communiquée par un intermédiaire; 9) l'observation des coutumes et pratiques commerciales; 10) l'existence de mécanismes d'assurance contre les messages non autorisés; 11) l'importance et la valeur de l'information contenue dans le message de données; 12) la disponibilité d'autres méthodes d'identification et le coût de leur mise en œuvre; 13) le degré d'acceptation ou de non-acceptation de la méthode d'identification dans le secteur ou domaine pertinent, tant au moment où la méthode a été convenue qu'à celui où le message de données a été communiqué; et 14) tout autre facteur pertinent. »

42. Ce modèle a été adopté par, entre autres, la société TenFour (www.tenfour.se) pour son logiciel de courrier sécurisé, par la société Atos (www.atos-group.com) pour son système de cartes Securicam et par la société PenOp (www.penop.com) pour l'intégration de son produit de signature biométrique aux PKIs.

nombre de certificats possibles.⁴³

Il semble clair que l'acte authentique se trouve au seuil d'une mutation profonde et inévitable, fruit de la progression inlassable des technologies de l'information et de la communication au sein des administrations. Cependant, cette mutation est principalement conceptualisée selon la nature de l'outil qui domine actuellement le marché. Cette domination, fruit de phénomènes conjecturels, impose une analyse extrêmement réductrice de la sécurité informatique, analyse qui subit une critique grandissante de la part des experts en sécurité.⁴⁴ Il est donc impératif que **la réflexion se poursuive sur les bases d'une analyse des besoins** et dans l'esprit d'une adéquation des technologies à ces besoins. C'est seulement au prix de cet effort que la spécificité de l'acte authentique comme outil de sécurisation juridique de droit civil pourra être maintenue, et non diluée par l'utilisation irréfléchie de technologies difficilement compatibles avec son principe.

43. Peut-être ceci peut-il fournir une motivation supplémentaire à la création d'un service public de la certification?

44. En fait, tout ceux qui ont osé se confronter au terrain : j'ai cité tout au long de ce document Ross Anderson, Bruce Schneier, Dorothy Denning.

Chapitre 4

Archivage

L'archivage des documents électroniques présente en soi des défis extraordinaires, et des problèmes inédits : comment, en effet, s'assurer de la pérennité des documents, quand toute l'infrastructure se renouvelle ? L'archivage et l'exploitation des actes authentiques introduit une dimension supplémentaire au problème : comment assurer la pérennité des technologies de signature électronique utilisées sur les documents archivés et/ou exploités ? L'interaction entre les multiples technologies utilisées est d'une complexité extraordinaire mais la solution réside peut-être tout simplement dans l'utilisation de mécanismes sociaux avec lesquels nous sommes déjà familiers. Les représentants de la Direction des Archives de France ont suggéré que « [l]a dématérialisation des documents, de leur forme de conservation à long terme et celle à venir de leur communication, est ... susceptible de remettre profondément en cause la structure institutionnelle de la politique d'archivage en France ... ».¹ Ce chapitre suggère que cette dématérialisation est également susceptible d'élargir les missions dont sont investies les institutions d'archivage en France, en leur adjoignant des fonctions supplémentaires de contrôle de l'intégrité.²

La première observation qui devrait guider toute réflexion sur l'archivage est que, « dans l'état actuel des choses, il est impossible de garantir la longévité et l'intelligibilité de données numériques pour même une seule génération humaine. »³ Ceci n'est pas une prédiction négative, mais simplement le constat lucide de l'immaturation de nos connaissances dans ce domaine, constat d'une communauté — celle des bibliothécaires et des archivistes — qui travaille depuis plusieurs années déjà sur la question. Cette communauté est une excellente source d'informations, ayant mené depuis de nombreuses années déjà une réflexion soutenue sur ces questions difficiles, de même que certaines expériences concrètes.

On peut distinguer trois problématiques distinctes concernant l'archivage des écrits électro-

1. CLEYET-MICHAUD, DHÉRENT, ERMISSE, « Remarques de la Direction des Archives de France sur la dématérialisation des actes authentiques », janvier 2001.

2. Fonction qu'elles exercent peut-être déjà implicitement : « Les auteurs classiques du droit de l'Ancien Régime (Pothier, Dumoulin) allaient jusqu'à admettre que la présence d'un document dans les archives publiques lui garantissait *ipso facto* un caractère d'authenticité. La jurisprudence actuelle n'irait sans doute pas aussi loin dans cette présomption ; elle n'en continue pas moins à accorder, en matière de publicité, une place privilégiée à l'entrée dans les fonds publics. » Hervé BASTIEN, *Droit des archives*, Paris : La Documentation Française, 1996, p. 7.

3. G. LAWRENCE, W. KEHOE, O. RIEGER, W. WALTERS, et Anne KENNEY, *Risk management of digital information: a file format investigation*, Council on Library and Information Resources, juin 2000.

niques : 1° la pérennité du support de l'écrit ; 2° la pérennité du format d'encodage de l'écrit ; 3° la pérennité des technologies de sécurisation de l'écrit. On connaît pas mal de choses sur le premier, beaucoup moins sur le second, presque rien sur le troisième, et **absolument rien sur l'interaction entre les trois.**

(1) Pérennité du support : Le premier sujet a été traité de main de maître par M. Dominique PONSOT, dans son rapport sur les technologies d'archivage numérique ou par microfilm.⁴ Je ne reviendrai donc pas sur cet aspect du problème, sauf lorsque nécessaire.

(2) Pérennité de l'encodage : Cinq solutions sont possibles pour la préservation de documents électroniques : 1° l'approche « copie papier » — imprimer une copie du document sur le papier ; 2° l'approche « standard universel » — développer un format standard et y migrer tous les documents ; 3° l'approche « musée de l'informatique » — conserver tout les équipements nécessaires à la lisibilité des formats ; 4° l'approche « émulation » — émuler par des logiciels les équipements périmés ; 5° l'approche « migration » — migrer périodiquement les fichiers vers les nouvelles versions des formats. Chacune de ces solutions comportent des avantages et des désavantages,⁵ mais la solution de migration périodique des fichiers semble la plus réaliste en ce moment.⁶

Pérennité de la sécurisation : Je ne connais *aucune* étude à ce jour qui puisse prétendre avoir sérieusement examiné la question de l'interaction entre la migration des documents et les procédés de sécurisation de ces documents. Très peu d'études *concrètes* existent sur la migration à grande échelle de bibliothèques de documents électroniques,⁷ et aucune sur les difficultés que présentent la migration simultanée des procédés mathématiques — qu'ils soient biométriques, cryptographiques ou basés sur le tatouage — vers les nouveaux formats d'encodage et les nouveaux supports.⁸ Or, en matière de sécurité électronique, ce sont les interactions imprévues entre les différents systèmes techniques qui procurent la plupart des failles pouvant être exploitées dans un but de fraude. On ne peut donc considérer les questions de migration de support, d'encodage et de sécurisation des écrits électroniques comme si elles existaient en isolation les unes des autres.⁹

Une solution à la pérennisation des procédés cryptographiques de signature est couramment mentionnée — la *resignature* — sans que l'on sache pourtant de quoi il en ressort exactement. Nous allons donc consacrer un peu d'énergie à explorer la logique de cette solution, en utilisant deux

4. M. Dominique PONSOT, *Valeur juridique des documents conservés sur support photographique ou numérique*, Observatoire juridique des technologies de l'information, 1995, disponible auprès des services de documentation du Premier ministre.

5. D. BEARMAN, « Reality and Chimeras in the Preservation of Electronic Records », *D-Lib Magazine*, April 1999.

6. La conservation d'une copie papier est certainement une mesure de sécurisation louable dans le contexte d'une familiarisation graduelle avec le document électronique, ou encore, la conservation simultanée de formats numériques et photographiques, l'approche adoptée par le CNAV pour son programme d'archivage des dossiers de retraite.

7. Une des plus intéressantes : supra, *Risk management of digital information: a file format investigation*.

8. Une étude produite dans le cadre du projet EESSI examine la question de la pérennité des signatures cryptographiques — voir Olivier LIBON, Andreas MITRAKAS, Angelika SCHREIBER, Jos DUMORTIER, Patrick VAN EECKE et Sofie VAN DEN EYNDE, « Trusted Archival Services », EESSI Report, août 2000

9. Par exemple, les fonctions de hachage qui fondent la signature cryptographique permettent de déceler toute modification au document numérique, fut-elle d'un seul bit. Cependant, elles n'effectuent aucune distinction entre une modification qui surviendrait à la suite d'une tentative de fraude, ou à la suite d'une migration du format de fichier.

Comme nous l'avons déjà vu (section 3.1.5), la cryptographie fonde et mesure sa sécurité sur la taille des clés : pour rendre un système cryptographique plus sécuritaire, il suffit d'augmenter la taille de la clé utilisée (avec évidemment un coût afférent sur la rapidité du système). Mais quelle taille faut-il utiliser et pour combien de temps procure-t-elle une sécurité suffisante ? La recherche d'une réponse à ces questions a donné naissance à une petite industrie académique parallèle : l'amélioration des algorithmes qui permettent de casser les procédés de cryptographie à clé publique — qu'ils soient fondés sur le problème de RSA¹³ ou d'autres approches, comme celle des courbes elliptiques.¹⁴

Ainsi, la sécurité procurée par la cryptographie diminue avec le temps, et il est loin d'être simple d'évaluer avec précision le *taux de cette diminution*.¹⁵ Un message pouvant être chiffré en toute confiance avec une clé de n bits au temps t pourra être déchiffré au temps $t + x$ années — et similairement pour un message signé cryptographiquement. Ceci est problématique car si l'écart entre les moments (2) et (4) est suffisamment grand, alors la vérification devant le juge ne pourra être valide, car l'assise fondamentale de la sécurité cryptographique sera faussée par la possibilité que la taille des clés initialement utilisée soit désormais insuffisante. C'est donc à ce problème que tente de répondre la resignature — resignature qui n'en est pas tout à fait une, comme nous allons à présent le constater.

4.2 L'approche EESSI

L'EESSI est un projet financé par la Communauté européenne dans le but de fournir des standards techniques réalisant les concepts de signature électronique énoncés par la Directive européenne.¹⁶ L'EESSI a déjà produit un certain nombre de documents qui nous permettront de mieux comprendre la logique qui sous-tend à l'archivage des signatures cryptographiques. Ces documents d'une nature extrêmement technique ne sont pas d'un accès facile, même pour des experts chevronnés. La lecture qui s'en suit se veut donc une visite guidée d'une petite partie de cet univers, celle qui se propose de résoudre le problème de l'archivage à long terme des signatures électroniques.

Dans l'optique de la signature cryptographique, la *vérification* de la signature est le seul moyen de s'assurer qu'une signature est valide, c'est-à-dire que l'on doit obtenir toutes les composantes qui forment la signature cryptographique et les fournir à l'algorithme de validation pour obtenir la réponse valide/invalid. La figure 4.2 reproduit le processus de validation du format de signature *ES-A* (*Electronic Signature – Archive Validation Data*), un format développé pour permettre la vérification de la signature longtemps après sa création. Dans le contexte de la ligne de vie d'une

13. Le dernier record est la factorisation d'un nombre de RSA de 512 bits — voir CAVALAR, DODSON, LENSTRA, et al., « Factorization of a 512-Bit RSA Modulus » in *Advances in Cryptology — EUROCRYPT 2000*, LNCS 1807, Springer 2000, pp. 1-17.

14. Voir, par exemple, <http://www.inria.fr/Presse/pre67-fra.html>.

15. Lors de la présentation du système RSA dans les pages du *Scientific American* d'août 1977, les auteurs offrirent un prix de 100 dollars à quiconque réussirait à déchiffrer un message chiffré à l'aide d'une clé de 425 bits, prédisant qu'un tel exploit nécessiterait des milliards d'années de calcul sur ordinateur. Or, le contenu de ce message (« *and the magic words are squeamish ossifrage* ») fut déchiffré le 27 avril 1994, moins de 20 ans après la publication du défi — voir <http://www.math.okstate.edu/wrightd/numthry/rsa129.html>.

16. Voir <http://www.ict.etsi.org/eessi/EESSI-homepage.htm>.

signature électronique (figure 4.1), le processus représenté à la figure 4.2 correspond à la validation effectuée au moment (2), dans le but de former un objet qui puisse être utilisé dans une validation au moment (4).

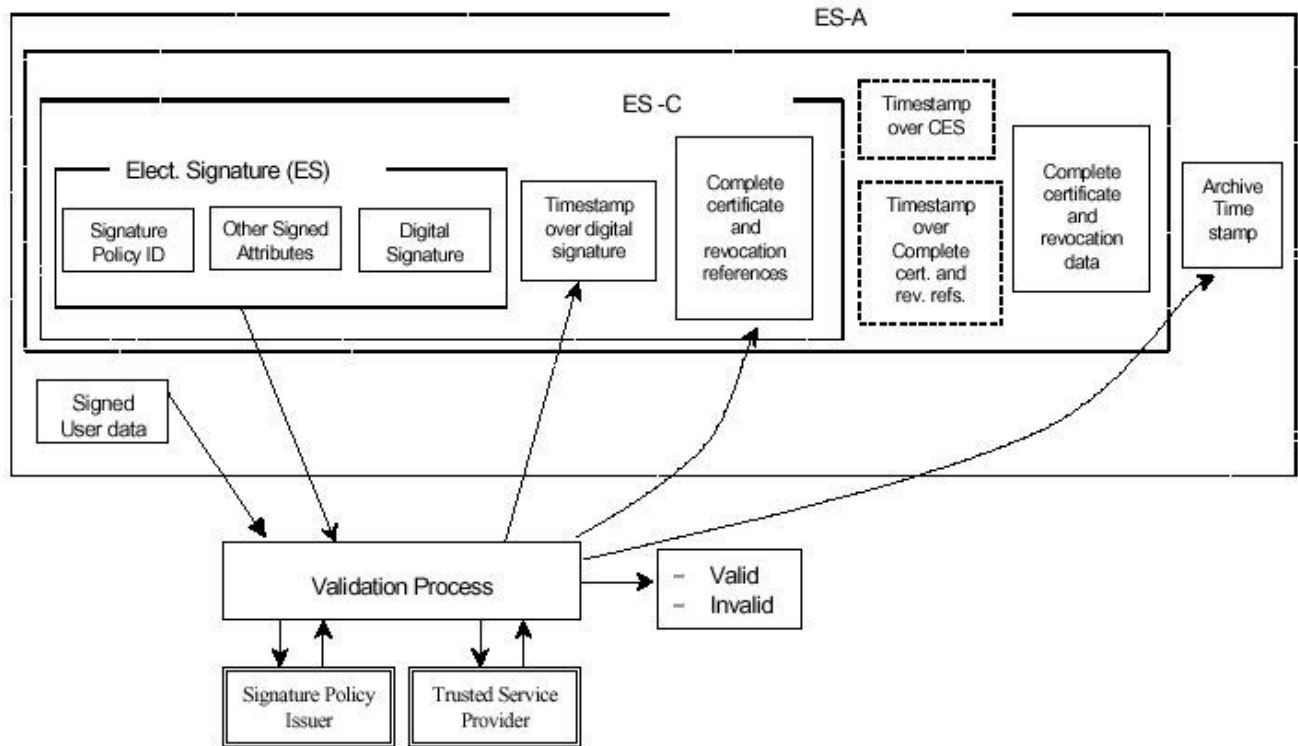


Figure 12: Illustration of an ES with Archive Validation Data

FIG. 4.2 – Validation du format de signature électronique ES-A — tiré de « *Electronic signatures formats* » ETSI TS 101 733 V1.2.2 (2000-12).

Le processus de validation implique six éléments principaux : (1) le rectangle intérieur marqué *ES*; (2) un algorithme de validation (rectangle marqué *Validation Process*) — qui reçoit en entrée les éléments de (1) et interagit avec (3) l'émetteur de la politique de signature (*Signature Policy Issuer*) et (4) des fournisseurs de service de confiance (*Trusted Service Provider*) pour finalement émettre en sortie une réponse (5) : la signature est valide ou n'est pas valide ; si la signature est valide, alors le processus produit (6) les éléments du rectangle supérieur (marqué *ES-A*), c'est-à-dire l'ensemble des éléments de la signature archivée. Le processus technique est le suivant : tout d'abord, le processus réalise la validation initiale (moment (2) de la ligne de vie), ce qui implique, dans l'ordre, de :

- 1° Obtenir le certificat à clé publique du signataire d'une des Autorités de confiance (*Trusted Service Provider*) ; obtenir le certificat à clé publique de l'Autorité de certification qui a signé le certificat du signataire ; si nécessaire, obtenir le certificat à clé publique de l'Autorité de certification qui a signé le certificat de la première Autorité de certification ; répéter jusqu'à l'obtention d'un certificat-racine ;
- 2° Vérifier sur les listes de révocation relatives à chacun de ces certificats si le certificat a été

révoqué. Si oui, la signature est invalide ; si non, vérifier, pour chacun des certificats, s'il a été suspendu ; si oui, attendre jusqu'à la fin de la période de suspension du certificat et reprendre le processus à l'étape 1° ;

3° Vérifier chacune des signatures sur les certificats, dans l'ordre du chemin de certification, en partant du certificat-racine ; si l'une de ces signatures est invalide, alors la signature est invalide ;

4° Calculer, à l'aide de la fonction de hachage spécifiée dans la Politique de signature (*Signature policy*), le condensé du document qui est l'objet de la signature (*Signed User data*) ;

5° Appliquer la clé publique contenue dans le certificat à clé publique du signataire au condensé calculé à l'étape précédente ; si le résultat est identique à la signature numérique (*Digital Signature*, troisième élément du rectangle ES), la signature est valide, sinon elle est invalide ;

Ensuite, le processus de validation construit l'objet qui pourra servir au moment (4), ce qui implique, dans l'ordre, de :

1° Horodater la signature numérique (bloc marqué *digital signature*) de façon à lui donner une date, date qui permettra d'établir l'existence de la signature relative aux listes de révocations et à la période de validité des certificats ; ajouter les références complètes à tous les certificats utilisés de même qu'aux informations de révocation — l'objet résultant a pour nom *ES-C* ;

2° Horodater le bloc *ES-C* ; horodater les références aux certificats et aux informations de révocation ; ajouter les certificats eux-mêmes et les données de révocation ; ajouter le document original, objet de la signature ;

3° Horodater le tout — l'objet résultant est une signature de format *ES-A* ;

4° Répéter l'étape précédente à chaque fois que la taille des clés ou la force des algorithmes utilisés par l'Autorité d'horodatage n'est plus jugée suffisante.

Le document ne précise pas à quoi ressemblerait le processus de validation de l'objet *ES-A* — une omission que l'on ne peut que regretter — mais il faudra ajouter aux vérifications mentionnées ci-dessus la vérification de chacun des horodatages. En effet, dans le modèle EESSI, l'horodatage d'un document par une Autorité d'horodatage s'effectue en deux opérations simples : (a) ajouter au document une date et une heure obtenue d'une source fiable et (b), signer le tout avec la clé privée de l'Autorité. Ces horodatages sont donc en fait des signatures supplémentaires pour lesquelles on devra également vérifier la chaîne de certification.¹⁷

4.3 Conclusions

Nous sommes donc à présent en mesure d'apprécier que le principe de resignature proposé par l'EESSI pose comme principe que, dans le contexte de la ligne de vie de la signature, le moment (4) de vérification devant le juge doit correspondre en tout point au moment (2), c'est-à-dire que *le juge doit être en mesure de répéter la même expérience qui a eu lieu lors de la vérification initiale*. Un tel principe suppose qu'il dispose également, 30, 60, ou 100 ans plus tard, de réalisations des algorithmes *identiques* à ceux qui ont initialement créés et vérifiés la signature — alors que les plates-formes

17. Le cas présenté est le plus simple (et le plus inintéressant, au niveau contractuel) : un document signé par une seule personne. Il ne traite ni la questions des certificats d'attributs, ni celle des signatures multiples sur un même document, chacun de ces certificats et de ces signatures dépendant de leur propre chaîne de certification.

informatiques, systèmes d'exploitation et logiciels originalement utilisés auront tous disparus depuis longtemps. Si on a depuis longtemps identifié que le changement technologique entraînait des conséquences importantes sur la *lisibilité des écrits électroniques*, on n'a pas constaté que le problème se pose avec tout autant d'acuité pour la *reproductibilité de la preuve électronique*.

Il est donc possible que le désir de construire un objet qui survive à la fois à l'expérience du moment (2) et au moment (4) procède d'hypothèses irréalistes. Une autre voie est-elle possible? Oui, et elle est en fait discrètement évoquée par le document EESSI à la section 4.3 :

« *Quand il y a exigence d'une signature valide sur une long durée sans utilisation de l'horodatage, alors il y a nécessité d'un enregistrement sécurisé de la date de la vérification (de la signature numérique) associé à la signature elle-même.* »¹⁸ (*Electronic Signature Formats*, sec. 4.3)

C'est-à-dire qu'une autre façon de procéder est tout simplement d'archiver au moment (3) une *attestation constatant la vérification de la signature* à un certain moment. L'écart entre le moment (2) de la vérification initiale et le moment (3) de l'archivage étant nettement plus court que celui séparant (2) et (4), une similarité satisfaisante entre les deux expériences est nettement plus plausible et permettrait une nette réduction de la complexité de la signature archivée. Le moment le plus approprié pour cette validation reste à raffiner, mais on peut imaginer qu'il pourrait avoir lieu à la fin de la période d'utilisation courante, où un tri est effectué entre les documents destinés à être conservés et ceux destinés à l'élimination.¹⁹

Quelque soit la solution adoptée, la question de l'archivage des technologies de sécurisation des documents électroniques pose le problème de définir quel sera exactement la forme de l'objet probatoire que l'on désire conserver, et la nature de l'expérience qui sera à même de faire « dire » à cet objet l'événement dont il est le témoin fidèle, longtemps après sa création. Cette question ne fait actuellement l'objet d'aucun débat, si ce n'est au sein de documents techniques inaccessibles aux non-initiés, même si elle est l'une des fonctions essentielles des archives.²⁰

18. « *When there is a requirement for long term signatures without timestamping the digital signatures, then a secure record is needed of the time of verification in association with the electronic signature [...].* »

19. Loi no. 79-18 du 3 janvier 1979 sur les archives, article 4.

20. Comme note finale, il est amusant d'observer les procédures que le document EESSI *Electronic Signature Formats*, document consacré à la spécification de mécanismes visant à établir l'origine et l'intégrité d'écrits électroniques, utilise pour établir *sa propre intégrité*: à la page 2, il est précisé que « ce document peut être rendu disponible sous plus d'une version électronique ou imprimée. Dans l'éventualité de différences perçues ou réelles quant au contenu de ces versions, la version de référence est celle en format PDF (*Portable Document Format*). En cas de dispute, la référence sera l'impression sur les imprimantes d'ETSI de la version PDF conservée sur un disque réseau conservé au secrétariat d'ETSI. » La politique d'ETSI, institution de normalisation des nouvelles technologies, avale donc explicitement le choix d'un encodage « fixant » l'information, de l'archivage par une institution de confiance et ... de la copie papier comme témoins privilégiés de l'intégrité de l'information électronique.

Chapitre 5

Conclusions

La sécurité est toujours fonction d'une multiplicité de facteurs : Toute solution de sécurité qui mérite ce nom mêle étroitement technique, déontologie, formation, information, réglementation, et pénalisation. Chacun de ces facteurs est nécessaire au succès de la solution adoptée, et aucun ne fonctionne isolément des autres. Il faut considérer le *faisceau des règles* qui, collectivement, créent une solution de sécurité. Aucune technologie n'est en mesure d'adresser à elle seule l'ensemble des maillons de la chaîne.

Reconfiguration et non pas élimination du risque : L'euphorie qui entoure la nouvelle économie tend à présenter les technologies de signature électronique comme un *progrès* sur la signature traditionnelle. Il faut plutôt réfléchir en termes de nouveaux avantages, et de nouveaux risques. L'informatisation n'a pas éliminé le risque de fraude, bien évidemment. Plutôt, on fait dorénavant face à une nouvelle configuration du risque. Il faut donc tenter de déterminer à quoi correspond cette nouvelle donne, une tâche cruciale puisqu'elle détermine la sélection des technologies de sécurité appropriées à ce risque.¹

Mixité du papier et de l'électronique : Nous ne vivons pas dans un univers du tout électronique, bien au contraire, et ce, particulièrement dans les professions judiciaires. Plutôt que de partir d'une idéologie de l'électronique à tout prix, il vaut mieux adopter un point de vue pragmatique et tenir compte du fait que le papier et l'électronique s'entremêlent à de multiples points de passage, et ce, pour de nombreuses années encore. Moins esthétique que les fantasmes de virtualité totale, mais plus proche de la réalité.²

1. Une tâche cruciale et difficile, puisque, comme le souligne Dorothy Denning, experte en sécurité informatique : « Il est facile d'évoquer des scénarios tels que "la bourse s'effondre après que des hackers bidouillent avec les ordinateurs de Wall Street" ou encore "deux avions se heurtent après que des terroristes tripatouillent les systèmes de navigation". Il est beaucoup plus difficile d'évaluer si de tes scénarios sont plausibles ou non. La grande question est celle-ci : Est-ce que quelqu'un peut lancer une attaque avec des conséquences catastrophiques et, si c'est le cas, quelles sont les probabilités qu'un tel événement se produise ? En vérité, personne ne le sait. » Dorothy E. DENNING, *Information Warfare and Security*, Addison Wesley, 1999.

2. Voir à sujet l'article de deux spécialistes du domaine, Ziming LIU et David G. STORK « Is Paperless Really More? Rethinking the Role of Paper in the Digital Age » *Communications of the ACM* 43:11(94-97), qui suggère que le développement du document électronique va stimuler une synergie entre l'univers papier et l'univers électronique, plutôt que l'un supplante simplement l'autre.

Sécurisation des documents matériels : Il existe un très riche réservoir d'expérience dans la production de documents sécurisés — passeports, billets de banque, cartes de crédits, etc. Ces documents utilisent typiquement un combinaison de techniques pour faire échec à la fraude, plutôt que de se reposer sur une seule méthode : papiers et encre spéciaux, filigranes, hologrammes, polymères et adhésifs, etc. Puisque le papier et l'électronique continueront de co-exister, on peut tirer profit des expériences déjà acquises dans ce domaine. Ainsi, la sécurisation électronique — que ce soit par la cryptographie ou par d'autres méthodes — continuera de co-exister avec des méthodes de sécurisation matérielle.

Hybridité du document électronique : Il faut considérer le document électronique comme un objet dépendant à la fois de son format d'encodage, du logiciel de lecture, des périphériques de visualisation et d'impression, et du matériel sous-jacent. On ne peut considérer ces éléments de façon indépendante, comme on pouvait le faire dans le cas du papier.

Pluralité des moyens techniques : Une des caractéristiques du débat sur la signature électronique, c'est que la discussion est dominée par des solutions basées sur les technologies cryptographiques. Même les textes qui se réclament d'une approche « technologiquement neutre » — la Directive européenne notamment — sont en fait hantés par la cryptographie. Or, il existe plusieurs autres solutions — la signature biométrique par exemple — qui permettent de réaliser la signature électronique, chacune avec ses caractéristiques. Ce document veut aider les juristes à s'extirper de la pensée unique qui voudrait imposer le couple cryptographie-PKI comme la seule réalisation possible de la signature électronique.

Efficacité juridique : D'où provient l'efficacité juridique de l'acte authentique? On désire un acte qui soit sûr, c'est-à-dire, entre autres choses, qui ne soulève pas de contentieux au niveau de son intégrité. Il y a tout lieu de se poser la question de la relation entre efficacité technologique et efficacité juridique. Un acte authentique électronique où, par exemple, le consentement des parties est pauvrement constaté risque d'être peu efficace juridiquement — il faut attentivement examiner la relation entre signature électronique et engagement moral.³ Cette question de la *mise en scène* de l'acte authentique a été peu discutée et pourtant, elle semble névralgique.⁴ Comme l'a récemment fait remarquer Me Lambert, la mise en scène de l'acte authentique doit permettre d'exprimer l'autorité de l'État qui est délégué à l'officier public et qui, en droit français au moins, garantit l'authenticité. Cette question d'apparence secondaire va, en fait, au cœur du problème de la relation entre technologies et institutions en sécurité informatique. L'ergonomie des procédés de sécurité informatique commence à provoquer de plus en plus d'intérêt au sein de la communauté scientifique : certaines études ont suggéré que 95% des atteintes à la sécurité étaient le fait d'erreurs de

3. C'est-à-dire que d'appuyer sur un bouton de souris n'a pas la même force d'expression du consentement et de solennité que la signature manuscrite.

4. Comme le souligne Pierre LEGENDRE : « Faire tenir ce que nous appelons l'État exige les grands moyens théâtraux, et l'Occident rationaliste ne déroge pas à l'expérience de l'humanité en matière d'institutions : il faut y croire, comme on croit à sa propre image. À propos des constructions, les architectes antiques parlaient de fermeté (*firmitas*), au sens où un bâtiment non seulement doit tenir debout selon les lois de la physique, mais aussi doit avoir l'air de tenir debout; il a la force d'une image. Il en est ainsi de l'État : ce sont les rites qui le font exister pour en faire une image au regard des croyants que nous sommes. . . . Il faut de la colle pour que tienne un État. Il faut les ancêtres, les images nostalgiques et tout le saint-frusquin des mises en scène ; ça passe par les liturgies politiques, par l'architecture des lieux, par toutes les formes d'écriture nécessaires à la ritualisation du pouvoir. » Pierre LEGENDRE, *Miroir d'un nation — l'École nationale d'administration*, Milles et une nuits, 2000.

configuration dues à la pauvreté des interfaces. ⁵

5. Bien que la littérature soit encore assez mince, plusieurs études commencent à soulever les préjugés anti-utilisateurs qui sous-tendent souvent les procédés de sécurité informatique : on tente de domestiquer l'utilisateur à utiliser un procédé qui va contre son entendement, et lorsque celui-ci se réticent, on le décrit comme "le maillon faible", "l'incompétent", de la chaîne. Voir, par exemple, Anne Adams et Martina Angella Sasse, "Users are not the enemy: Why users compromise security mechanisms and how to take remedial measures", *Communications of the ACM* 42:(12), p. 40-46 ; Alma Whitten and J.D. Tygar, "Usability of security: A case study", Carnegie Mellon University Technical Report, décembre 1999.