Intégrité, signature et processus d'archivage

Françoise Banat-Berger[1] et Anne Canteaut[2]

Les textes législatifs concernant l'écrit électronique font de sa conservation une des conditions essentielles de sa valeur juridique, sans toutefois en définir précisément les modalités. La conservation des documents électroniques est donc un enjeu majeur pour qui souhaite s'assurer qu'ils constitueront une preuve recevable en justice, au même titre que les documents sur support papier, plusieurs années après leur établissement.

De quel objet doit-on garantir l'intégrité ?

Qu'est-ce qu'un écrit électronique ?

Le point de départ de notre travail est naturellement l'article 1316-1 du Code civil qui dispose que « l'écrit sous forme électronique peut être admis à titre de preuve au même titre que l'écrit sur support papier, sous réserve que puisse être dûment identifiée la personne dont il émane et qu'il soit établi et conservé dans des conditions de nature à en garantir l'intégrité ». Curieusement, lorsque nous avons entamé cette étude pluri-disciplinaire, nous pensions que nous allions principalement nous heurter à des différences de définition du terme « intégrité ». Au contraire, nous avons rapidement pris conscience du fait que les informaticiens et les archivistes avaient de l'intégrité des compréhensions assez proches : l'intégrité d'un objet est le fait qu'il n'ait subi aucune altération, qu'elle soit accidentelle ou intentionnelle. La définition de l'intégrité ne nous pose donc pas de problème ; ce qui nous pose problème est en fait de déterminer l'objet dont nous voulons garantir l'intégrité. Dans l'article 1316-1 C. civ., l'intégrité porte sur « l'écrit sous forme électronique ». Nous avons donc besoin d'une définition précise de ce concept ; celle-ci est fournie par l'article 1316 : « une suite de lettres, de caractères, de chiffres ou de tout autre signes ou symboles, dotés d'une signification intelligible, quels que soient leur support et leur modalité de transmission ». Entrent donc dans cette définition un écrit alphabétique, un message en morse, une séquence d'ADN et bien entendu un écrit numérique, c'est-à-dire un écrit représenté sous forme de nombres.
Pour un informaticien, un écrit numérique est une suite de symboles binaires, c'est-à-dire de 0 et de 1, puisqu'il s'agit de données créées, stockées, transmises, traitées ou affichées au moyen d'un système numérique et non analogique (au sens où l'on parle de « photographie numérique » par exemple). Ainsi, pour comprendre ce qu'est un écrit numérique, il est nécessaire de comprendre comment une telle suite de 0 et de 1 peut représenter l'information que l'on veut stocker ou transmettre.

Codage des caractères

Dans ce contexte, la première question naturelle est celle de la représentation d'un texte alphabétique

[1] Direction des Archives de France, département de l'innovation technologique et de la normalisation, francoise.banat-berger@culture.gouv.fr
[2] INRIA - projet CODES, B.P. 105, 78153 Le Chesnay cedex, Anne.Canteaut@inria.fr

sous forme binaire, c'est-à-dire par une suite de 0 et de 1. Il existe quantité de méthodes pour « coder » les caractères sous forme binaire. Le code Morse en est un exemple qui a l'avantage d'adapter la longueur de la suite correspondant à chaque caractère à sa fréquence d'apparition, même si cette fonctionnalité s'avère parfois peu pratique --- dans la plupart des applications usuelles, on préfère représenter chaque caractère par le même nombre de symboles binaires. L'unité élémentaire de la représentation binaire est le bit, abréviation de l'anglais « binary digit ». A l'aide d'une suite composée de n symboles binaires (un mot de n bits), on peut ainsi coder 2^n caractères. De nos jours, le codage des caractères alphabétiques (au sens large, c'est-à-dire en incluant les chiffres, les signes de ponctuation et autres caractères usuels présents sur un clavier) est effectué suivant des normes bien établies, mais le choix de la norme évolue naturellement avec la technologie. Une des plus anciennes qui est aussi la plus courante (et donc la plus largement inter-opérable) est le code ASCII (American Standard Code for Information Interchange). Il s'agit d'un codage des caractères usuels par des suites de 7 bits. Ce choix était en effet naturel aux débuts de l'informatique car les ordinateurs manipulaient alors généralement les données par blocs de 8 bits[3], et l'un de ces 8 bits ne pouvait pas être utilisé pour coder de l'information puisqu'il servait uniquement à détecter les erreurs de transmission[4]. Ce système de codage permet donc de représenter 128 caractères, ce qui suffit pour les caractères usuels présents sur un clavier américain.

Le tableau suivant donne la correspondance entre les caractères et les suites de 7 bits dans le code ASCII. Remarquons qu'une suite de bits est généralement identifiée à un entier par le biais de sa représentation en base 2. Par exemple, le caractère « C » est représenté par la suite de 7 bits 1000011 qui correspond à l'entier 67 puisque $67 = 1 * 2^6 + 0 * 2^5 + 0 * 2^4 + 0 * 2^3 + 0 * 2^2 + 1 * 2^1 + 1 * 2^0 = 64 + 2 + 1$. Les caractères 0 à 31 de cette table, ainsi que le caractère de code 127, ne sont pas imprimables : il s'agit de caractères de contrôle, par exemple, HT est la tabulation horizontale, BEL est le caractère qui provoque un signal sonore, ESC l'échappement…

Tableau 1: le codage des caractères par le code ASCII

	Code binaire	Code décimal		Code binaire	Code décimal		Code binaire	Code décimal		Code binaire	Code décimal
NUL	0000000	0	SP	0100000	32	@	1000000	64	`	1100000	96
SOH	0000001	1	!	0100001	33	A	1000001	65	a	1100001	97
STX	0000010	2	"	0100010	34	B	1000010	66	b	1100010	98
ETX	0000011	3	#	0100011	35	C	1000011	67	c	1100011	99
OT	0000100	4	$	0100100	36	D	1000100	68	d	1100100	100
ENQ	0000101	5	%	0100101	37	E	1000101	69	e	1100101	101
ACK	0000110	6	&	0100110	38	F	1000110	70	f	1100110	102
BEL	0000111	7	‘	0100111	39	G	1000111	71	g	1100111	103
BS	0001000	8	(	0101000	40	H	1001000	72	h	1101000	104
HT	0001001	9	)	0101001	41	I	1001001	73	i	1101001	105
LF	0001010	10	*	0101010	42	J	1001010	74	j	1101010	106
VT	0001011	11	+	0101011	43	K	1001011	75	k	1101011	107
FF	0001100	12	,	0101100	44	L	1001100	76	l	1101100	108
CR	0001101	13	-	0101101	45	M	1001101	77	m	1101101	109
SO	0001110	14	.	0101110	46	N	1001110	78	n	1101110	110
SI	0001111	15	/	0101111	47	O	1001111	79	o	1101111	111
DLE	0010000	16	0	0110000	48	P	1010000	80	p	1110000	112
DC1	0010001	17	1	0110001	49	Q	1010001	81	q	1110001	113
DC2	0010010	18	2	0110010	50	R	1010010	82	r	1110010	114

[3]Un mot de 8 bits est appelé un octet (Byte en anglais).

[4]Comme la clef associée au NIR (numéro de sécurité sociale) qui permet de détecter les erreurs de saisie.

DC3	0010011	19	3	0110011	51	S	1010011	83	s	1110011	115
DC4	0010100	20	4	0110100	52	T	1010100	84	t	1110100	116
NAK	0010101	21	5	0110101	53	U	1010101	85	u	1110101	117
SYN	0010110	22	6	0110110	54	V	1010110	86	v	1110110	118
ETB	0010111	23	7	0110111	55	W	1010111	87	w	1110111	119
CAN	0011000	24	8	0111000	56	X	1011000	88	x	1111000	120
EM	0011001	25	9	0111001	57	Y	1011001	89	y	1111001	121
SUB	0011010	26	:	0111010	58	Z	1011010	90	z	1111010	122
ESC	0011011	27	;	0111011	59	[	1011011	91	{	1111011	123
FS	0011100	28	<	0111100	60	\	1011100	92	|	1111100	124
GS	0011101	29	=	0111101	61	]	1011101	93	}	1111101	125
RS	0011110	30	>	0111110	62	^	1011110	94	~	1111110	126
US	0011111	31	?	0111111	63	_	1011111	95	DEL	1111111	127

Les 128 caractères du code ASCII ne suffisent cependant pas à représenter tous les caractères utilisés dans toutes les langues. Par exemple, les caractères accentués du français n'y figurent pas. Pour pallier ce manque, une norme internationale maintenant reconnue par la grande majorité des logiciels lui a succédé : la norme ISO-8859. Cette norme code les caractères par des suites de 8 bits, ce qui offre 256 possibilités. Les 128 premiers caractères, correspondant aux entiers entre 0 et 127, sont identiques à ceux de la norme ASCII, ce qui permet la compatibilité entre les deux normes. Les 128 possibilités supplémentaires offertes par la norme ISO-8859, les entiers entre 128 et 255, servent donc à coder les caractères spécifiques aux autres langues. Pour cela, la norme possède plusieurs déclinaisons, une par grand groupe linguistique. Ainsi, la langue française utilise la déclinaison ISO-8859-1, également appelée ISO-Latin-1, qui contient les caractères accentués, le « ç » (code 231) par exemple, mais aussi le « ß » ou le « å ». La variante ISO-8859-2 correspond, elle, aux langues d'Europe centrale et permet d'utiliser des caractères comme « c », la variante ISO-8859-5 inclut les caractères cyrilliques, ISO-8859-6 les caractères arabes, ISO-8859-7 les caractères grecs…
L'augmentation des débits de transmission a ensuite amené l'idée que l'on pouvait transmettre plus d'un octet pour chaque caractère. Ainsi, la norme Unicode étend ISO-8859, et fait correspondre 16 bits à un caractère, ce qui permet d'en représenter 65536, et donc de couvrir la plupart des langues (et même par exemple les idéogrammes du linéaire B, des notations musicales…[5]).

Dans ce contexte, un « éditeur de texte » (par exemple le « bloc-notes » sous Windows, Emacs sous Linux...) est un logiciel qui permet de visualiser le contenu d'un écrit numérique simplement en interprétant chaque suite de 7 bits par le caractère associé dans la norme ASCII (ou la norme ISO-8859). Autrement dit, l´écrit numérique correspondant à un texte non formaté (c'est-à-dire sans gras, italique, variation de polices de caractères...) est une suite de 0 et de 1 correspondant à la succession des codes ASCII (ou ISO-Latin-1, ou Unicode) des différents caractères du texte. Par exemple le fichier numérique que l'on visualise de la manière suivante avec un éditeur de texte est constitué de 272 octets (c'est-à-dire 2176 bits), puisqu'il comporte 272 caractères. La visualisation au moyen d'un éditeur de texte fournit donc une bonne idée de ce qu'est un document électronique, autrement dit de ce que contient le fichier numérique considéré.

[5] La liste des caractères Unicode est disponible par exemple sur http://www.unicode.org/fr/ .

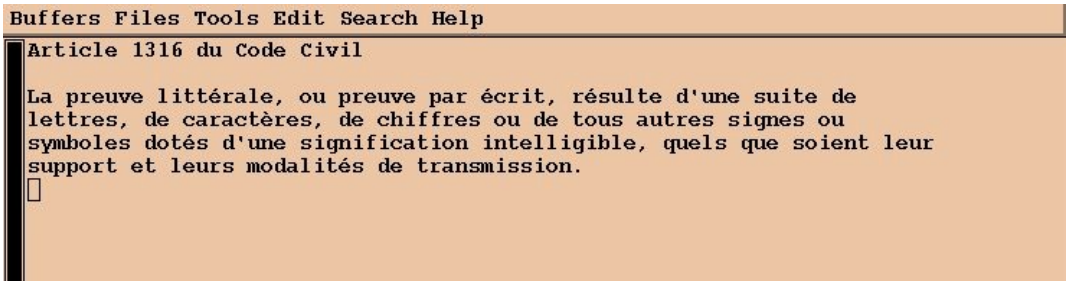

Figure 1 : visualisation d'un fichier texte avec un éditeur de texte

Format d'un document numérique

La particularité des fichiers créés au moyen d'un éditeur de texte comme celui que nous venons de voir est donc que le seul codage utilisé est celui des caractères. Lorsque la norme de codage des caractères est donnée, il y a donc correspondance univoque entre le texte et le document numérique. Le format d'un tel document est appelé texte brut ou parfois simplement ASCII. Le nom de ces fichiers est généralement caractérisé par le suffixe .txt ou .ascii.
A l'inverse, tous les autres types de fichiers utilisent un codage supplémentaire (pour spécifier la police de caractères, ou décrire des objets beaucoup plus complexes comme une image par exemple). La correspondance entre le document numérique et son contenu perceptible, c'est-à-dire sa visualisation au moyen de divers logiciels, est alors beaucoup plus compliquée et n'est plus univoque.

Par exemple, le format HTML permet de spécifier des propriétés sur les polices de caractères, de faire des tableaux... au moyen de balises. Ainsi, le document que l'on visualise au moyen d'un navigateur sous la forme présentée ci-dessous contient les codes ASCII des différents caractères, mais également des instructions de formatage. Il est formé de 514 octets (en comparaison des 272 octets du texte brut) et sa visualisation par un éditeur de texte fournit le contenu du document numérique correspondant.

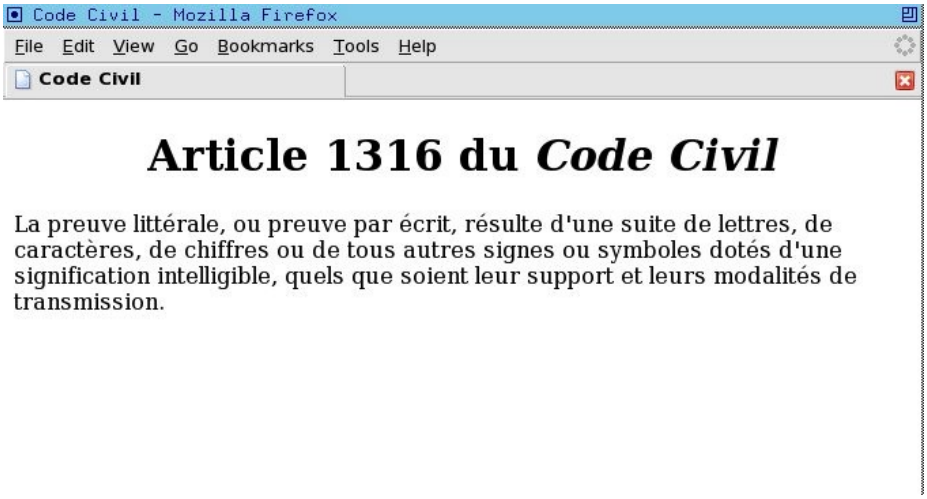

Figure 2 : Visualisation d'un document HTML avec un navigateur

```
Buffers Files Tools Edit Search Help
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.o\
rg/TR/html4/loose.dtd">
<html>
<head>
<title>Code Civil</title>
</head>

<body bgcolor="#FFFFFF">
<h1><center><B>Article 1316 du <i>Code Civil</i></B></center></h1>

<p>La preuve littérale, ou preuve par écrit, résulte d'une suite de
lettres, de caractères, de chiffres ou de tous autres signes ou
symboles dotés d'une signification intelligible, quels que soient leur
support et leurs modalités de transmission.
</p>

</body>
</html>
```

Figure 3 : visualisation du document HTML précédent avec un éditeur de texte

Considérons maintenant un document numérique au format JPEG. Ce document est beaucoup plus gros, puisqu'il est constitué de 56959 octets. La visualisation du début de ce fichier avec un éditeur de texte nous fournit un document incompréhensible pour un humain. Par contre, sa visualisation par un logiciel d'affichage d'image nous donne un document dont le contenu informationnel est très proche de celui des documents précédents.

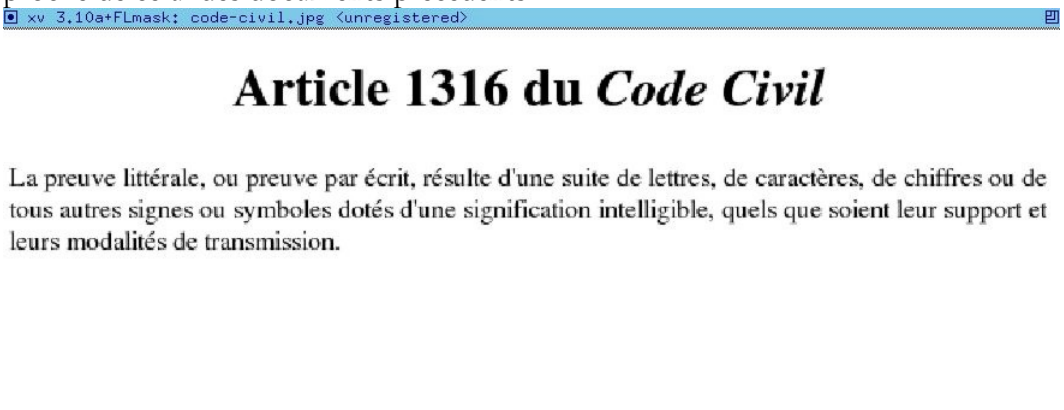

Figure 4 : visualisation d'un document JPEG avec un logiciel d'affichage d'images

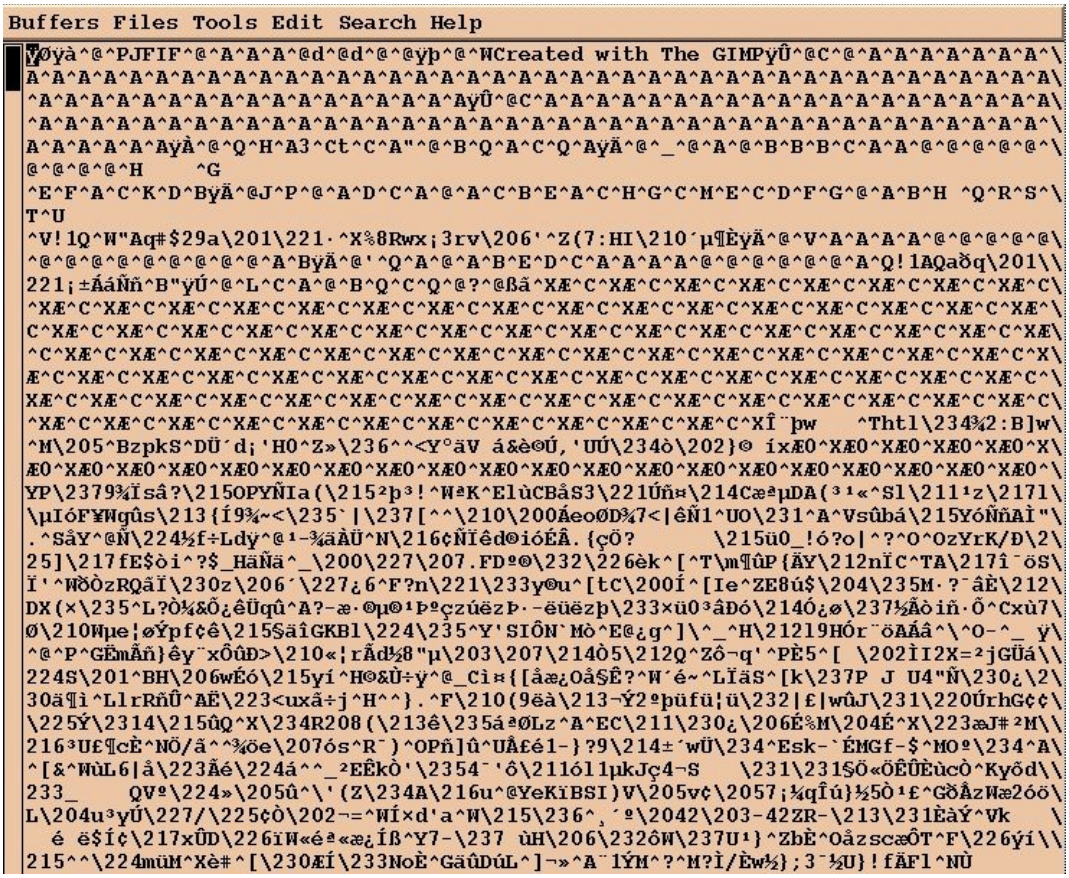

Figure 5 : visualisation du début du même document JPEG avec un éditeur de texte

De même, un document au format Word dont le contenu informationnel est identique aux précédents correspond à une suite de 8192 octets (c'est-à-dire 30 fois plus longue que le document au format texte). Son contenu observé au moyen d'un éditeur de texte montre que le texte est accompagné de nombreuses indications dont la plupart sont incompréhensibles sans l'aide du logiciel Word, mais parmi lesquelles on peut aussi reconnaître le nom de l'utilisateur et le numéro de licence des différentes versions du logiciel Word qui ont servi à créer le document.

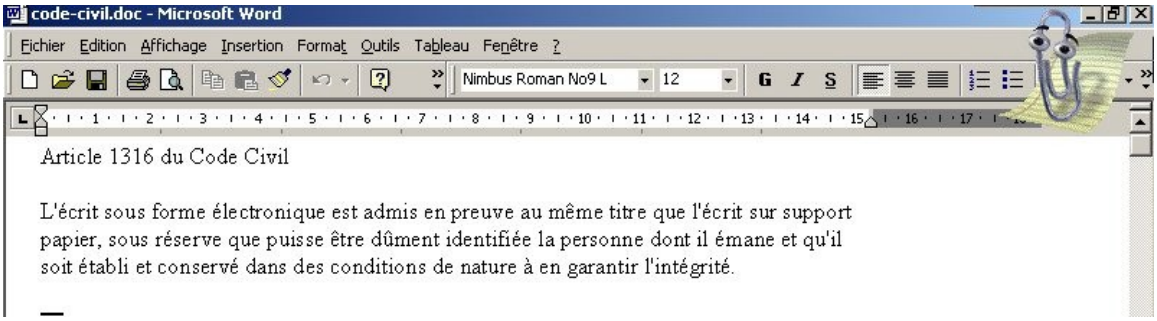

Figure 6 : visualisation d'un document au format .doc avec le logiciel Word

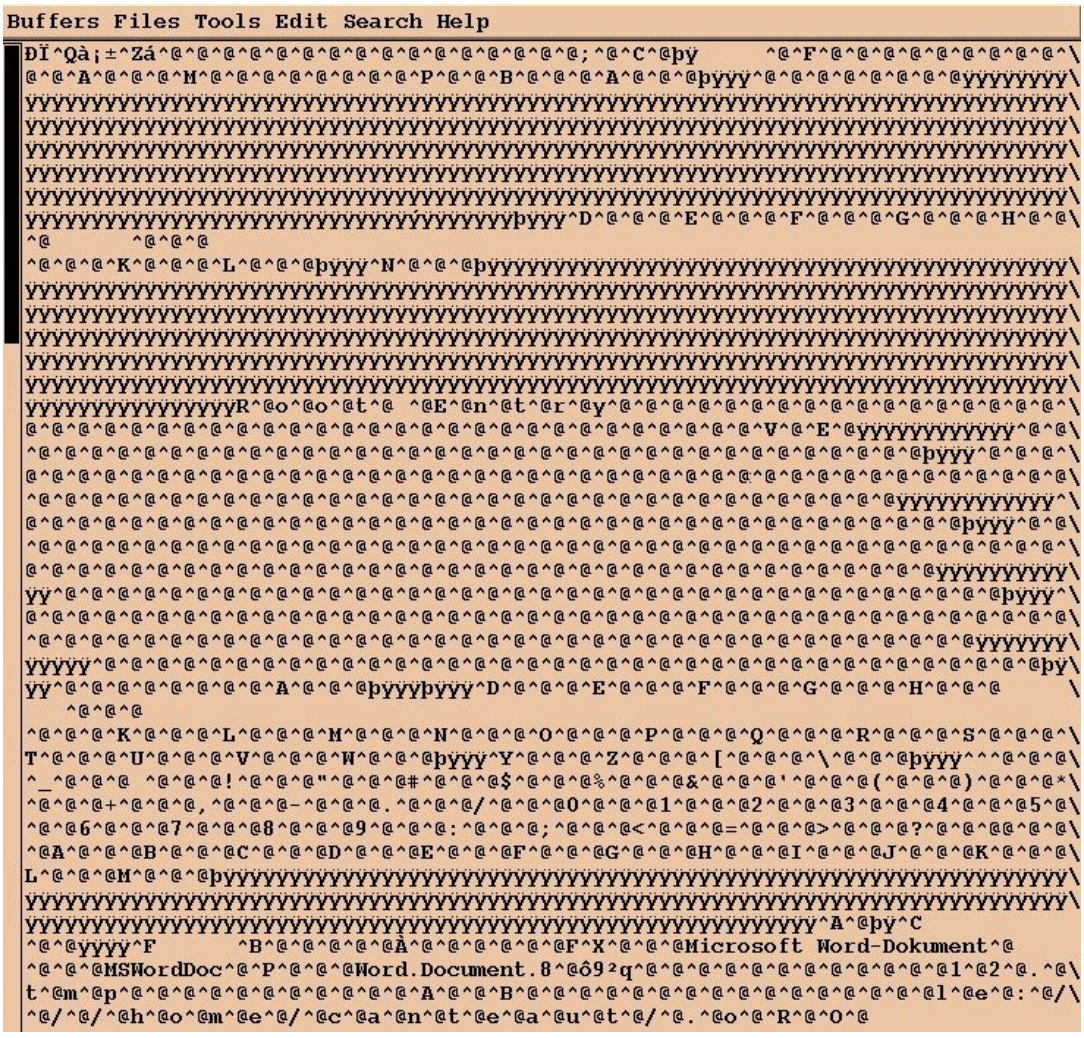

Figure 7 : visualisation du début du même document .doc avec un éditeur de texte

Un dernier exemple de format complexe est celui du document dont le début est visualisé par l'éditeur sous la forme suivante.

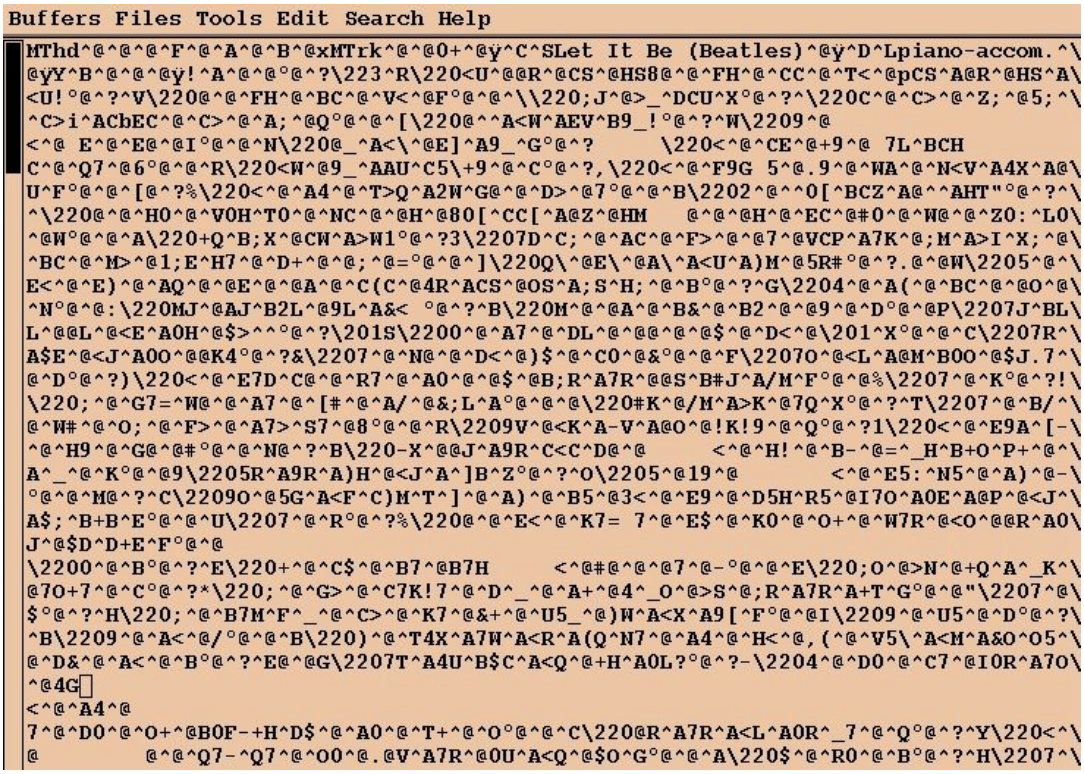

Figure 8 : visualisation d'un document sonore avec un éditeur de texte

Il s'agit cette fois-ci d'un document correspondant à un morceau de musique (dans un format audio) dont il apparaît clairement qu'il répond également à la définition d'écrit électronique de l'article 1316-1 du Code Civil, définition qui inclut absolument tous les documents numériques quel que soit leur format et leur nature perceptive (texte, image, son...).

A travers tous ces exemples, on voit clairement qu'un même contenu informationnel peut être exprimé par des documents numériques extrêmement différents (dont la taille peut varier par exemple d'un facteur 200). Pourtant, c'est bien le document numérique sous forme d'une suite de 0 et de 1 qui caractérise cet écrit, puisqu'il s'agit de la seule représentation tangible dont on dispose. C'est donc de cette suite binaire, et d'elle seule, dont nous pouvons techniquement garantir l'intégrité et c'est également à elle que l'on applique la signature. Par ailleurs, si certains formats sont simples et peuvent être interprétés par un humain, d'autres requièrent des matériels ou des logiciels sophistiqués pour que l'on puisse accéder à leur contenu. Or, ces derniers n'échappent pas aux évolutions technologiques, ce qui rend les documents que nous avons mentionnés impossibles à lire au bout d'un certain temps.

Une nouvelle définition de l'intégrité

Garantir la lisibilité d'un document archivé

L'intégrité est évidemment une notion qui paraissait centrale pour les rédacteurs de l'article 1316-1 du Code civil parce que dans l'environnement numérique, l'information est très facilement falsifiable.

S'il y a une faille de sécurité dans un système, des milliers de documents vont ainsi pouvoir être modifiés en une seule fois. C'est la raison pour laquelle l'intégrité est apparue comme une notion centrale dans le processus.
Cependant, affirmer qu'il convient d'établir et de conserver un acte de manière à en garantir l'intégrité, cela pourrait être interprété comme l'obligation impérieuse de considérer un document numérique comme un objet physique qui une fois établi, doit être conservé comme tel. Malheureusement, ce n'est pas ainsi que les choses se passent. En effet, comme le montrent les exemples précédents, l'affichage d'une information numérique est le résultat d'une suite complexe de traitements intervenant dans différentes couches de l'ordinateur (couches physiques, de structure, sémantiques...). C'est par le biais d'une superposition de matériels, de logiciels, de systèmes d'exploitation, de périphériques, que à un moment donné, pourra être affiché de par leur interaction harmonieuse, un écrit. Or, ces différents composants évoluent à des rythmes différents (obsolescence technique de plus en plus rapide). Par conséquent, conserver une information numérique et permettre sa représentation dans un laps de temps X, c'est conserver la capacité à la représenter dans l'environnement matériel et logiciel qui sera alors utilisé. Ainsi si on prétend conserver un document sous un format bureautique donné qui est lié à un système d'exploitation précis, le temps de la conservation ne pourra excéder la durée de vie de ce système d'exploitation. Pour pouvoir représenter un même document des années après, il va falloir retrouver une harmonie entre les matériels, les logiciels et les périphériques qui étaient utilisés à l'époque de création du document. Il convient par conséquent de rendre l'information à conserver la plus indépendante possible de son environnement technologique et il conviendra de procéder à des migrations notamment des migrations de format (par exemple d'un format dit fermé à un format dont les spécifications sont publiques, de manière à rendre possible de pouvoir si nécessaire, ré-écrire quand on le souhaitera, le programme qui permettra d'interpréter correctement l'information). Or, toute migration va affecter la suite de 0 et de 1 qui constituent l'information à sa base et donc rendre inopérant le concept d'objet physique inaltéré.

Cette question, si elle n'a pas été abordée au moment de l'élaboration de la loi du 13 mars 2000 et de ses décrets d'application, du point de vue de la conservation pérenne des informations numériques, commence à être posée depuis un certain temps, et tout naturellement au moment de l'élaboration des décrets concernant les minutes électroniques des notaires et huissiers, ce type de document étant réputé être conservé indéfiniment. C'est notamment la raison pour laquelle dans les décrets qui ont été finalement été élaborés, un nouveau concept juridique apparaît, à savoir que les migrations nécessaires à assurer la lisibilité du document, ne lui retirent pas son caractère d'original. On a par ailleurs réfléchi à cette problématique dans le cadre du groupe de travail sur la conservation des documents électroniques, constitué au sein du forum des droits sur l'Internet (recommandation sur le sujet publiée en décembre 2005).
Ainsi, selon la recommandation, trois critères cumulatifs sont nécessaires pour garantir l'intégrité :

- la lisibilité du document,
- la stabilité de son contenu informationnel,
- la traçabilité des opérations sur ce document.

La lisibilité désigne « la possibilité d'avoir accès au moment de la restitution du document à l'ensemble des informations qu'il comporte ». C'est là qu'intervient un concept central pour les professionnels de l'information, qu'ils soient archivistes, documentalistes, bibliothécaires : les métadonnées, que l'on peut comparer schématiquement à ce qu'est pour un médicament, sa notice. Il s'agit en effet des données sur les données qui permettent d'interpréter correctement l'information. Si on conserve les données et les métadonnées, ces dernières vont pouvoir redonner un sens à ce que l'on est censé pérenniser.

Prenons comme exemple, celui d' un export d'une base de données de l'administration. Les données sont exportées dans un format dit à plat. Elles sont lisibles : on lit en effet des suites de chiffres, numéros, caractères, mots séparés par des points-virgules. Si on conserve simplement ces données ainsi, l'archiviste ne pourra pas restituer une information intelligible à la personne autre que celle qui l'a versée aux archives. En effet, celui qui a versé sait quelle est la structure de sa base de données. Il aura donc la base de connaissance suffisante pour pouvoir comprendre l'information que l'archiviste va lui restituer. La donnée pour devenir une information porteuse de sens, devra être accompagnée d'un certain nombre d'informations comme la structure de la base de données (nature et définition des différentes tables qui la composent, dictionnaire des données, longueur des champs). Cela ne sera pas encore suffisant : il faudra également verser à l'archiviste toutes les nomenclatures à partir desquelles certains champs sont codés (codes et signification de leurs libellés). Ces métadonnées constituent ce que l'on appelle des métadonnées descriptives, de contenu. Outre ces métadonnées de contenu, on trouve d'autres types de métadonnées :

- métadonnées contextuelles : provenance du document, histoire du document ;
- métadonnées de gestion : date du versement dans le système des archives, qui s'en est occupé, ce qui est arrivé après à ce document ;
- métadonnées techniques : qui servent précisément à la pérennisation (notamment format des données, format d'encodage des caractères, type de compression...). Ce type d'information est capitale. En effet, elle va permettre d'identifier précisément et de pouvoir programmer si nécessaire sa conversion vers des formats plus pérennes. L'idéal est que ce type d'informations se trouve rassemblé, structuré, codifié dans ce qu'on appelle des registres de formats, l'objectif étant que les systèmes d'archivage puissent alors s'interfacer avec de tels registres : ceux-ci permettent en effet de disposer d'identifiants uniques par type de format et par version de format, de contrôler ainsi que ce qui est versé est bien le format annoncé (le fait de disposer des extensions des fichiers n'apportent aucune information fiable sur les formats dans lesquels ces derniers ont été encodés), et ainsi de les gérer, les convertir, etc. Il en existe malheureusement encore très peu à l'échelle d'un pays ou d'une communauté. Un exemple en est le code MIME, qui est relativement simple, est compris par tous les systèmes mais qui est trop générique et ne permet que retrouver de grandes catégories de familles de formats.

Ces métadonnées sont un élément central de la norme OAIS[6] (modèle conceptuel d'une archive). L'OAIS tend à caractériser ce que l'on appelle les objets d'information, les métadonnées nécessaires à leur préservation et l'organisation nécessaire pour les collecter, conserver et communiquer dans le cadre d'un système d'archives.
L'objet contenu de données et l'information de représentation constituent un contenu d'information qui est ce que l'on cherche à pérenniser autant que nécessaire (ainsi une information de représentation d'un document encodé en PDF. sera l'indication précise de la version de PDF avec laquelle il est représenté). Ce contenu d'information devra être accompagnée de ce que l'on appelle des informations de pérennisation (catégorisées en informations de provenance, identification, contexte, intégrité). L'ensemble constitue un paquet d'information, accompagné de ses informations

[6] La norme ISO 14721:2003 (Systèmes de transfert des informations et données spatiales -- Système ouvert d'archivage de l'information -- Modèle de référence), plus connue sous le nom de modèle OAIS (Open Archival Information System) est consultable à l'adresse suivante :
http://www.ccsds.org/CCSDS/documents/650x0b1.pdf.
Une traduction française, en cours de normalisation, est accessible à l'adresse suivante :
http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf.

de paquetage (la manière dont ces éléments sont ordonnés). A partir de ces métadonnées, des informations seront automatiquement extraites pour alimenter le système de gestion documentaire qui permettra de retrouver les documents.

Prenons l'exemple d'un message électronique : le contenu de données est la suite de bits qui constituent le message et les pièces jointes ; les informations de représentation sont des informations sur le code des caractères (ASCII, Unicode, etc.), l'encodage des pièces jointes (encodage base 64 pour les messages électroniques), le format du message et des pièces jointes ; les informations de pérennisation sont l'identité du destinataire, de l'expéditeur, la présence d'une signature électronique.
Pour une base de données, le contenu de données est la suite de bits correspondant aux différentes tables ; les informations de représentation sont celles relatives à la structure de la table, aux nomenclatures, au dictionnaire de données, aux codes de caractères, aux liens entre les tables ; les informations de pérennisation répondent quant à elles aux questions suivantes : à quoi servait cette base ? Pour qui était-elle faite ? Combien de temps a-t-elle fonctionné ? Quelles étaient les mises à jour ? etc.

De manière générale, seule l'utilisation de formats de données ouverts permet de garantir la lisibilité d'un document numérique au cours du temps. Les notions de standard et de format ouverts sont définies par la loi pour la confiance dans l'économie numérique du 21 juin 2004, dans son article 4 : « On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en oeuvre. » Cette catégorie de formats s'oppose donc à celle des formats fermés, qui sont des modes de représentation dont les conventions sont tenues secrètes ou dont l'utilisation est soumise à des droits (protection par des brevets…). Le format `doc` de Word est ainsi un exemple de format fermé. Dans le contexte de l'archivage, l'immense avantage des formats ouverts réside donc dans le fait que chacun a la possibilité, à tout moment, de produire un outil permettant d'afficher et de manipuler les données représentées sous ce format.
Cette caractéristique est naturellement essentielle pour garantir la lisibilité des documents, mais elle intervient également au moment de la signature du document. En effet, apposer une signature à un document numérique représenté par un format fermé (dont les spécifications ne sont pas publiques) implique donc de ne pas avoir accès au contenu du document numérique que l'on signe. Ainsi, si l'on signe un document au format doc, on signe non seulement le contenu du document tel qu'il apparaît à l'écran, interprété par le logiciel Word, mais on signe également le numéro de licence du logiciel utilisé pour le produire, des informations sur la machine sur laquelle il a été créé, parfois même l'historique du document. Ceci nous conduit donc nous interroger sur le statut juridique de la signature d'un tel document dans la mesure où la signature est une manifestation de consentement alors que le signataire n'a pas pleinement conscience du contenu du document qu'il signe.

Pour prendre la mesure du problème, rappelons que des informations importantes ont été révélées à l'insu de leurs auteurs par la simple diffusion de documents numériques sous un format fermé. Un exemple célèbre est celui d'un document émanant du gouvernement britannique sur l'existence d'armes de destruction massive en Irak et publié sur le site web du 10 Downing Street en février 2003 au format Word[7]. Ce document a par la suite fait l'objet d'une enquête pour cause de plagiat. Or, l'examen de la totalité du contenu de ce document fournit naturellement des informations supplémentaires, dont l'existence n'est pas documentée. Ainsi, si l'on ouvre ce fichier à l'aide d'un éditeur de texte, on constate qu'une partie du document contient (sous forme de caractères séparés par des « ^@ ») des renseignements sur les personnes ayant effectué les sauvegardes successives de

[7] Le fichier Word incriminé est disponible sur http://www.computerbytesman.com/privacy/blair.doc .

ce document, ce qui permet de reconstituer son cheminement. On peut par exemple lire à la première ligne de la figure 9: « JPratt A:\Iraq - security.doc » puis « ablackshaw C:\ABlackshaw\Iraq - security.doc », montrant que John Pratt a sauvegardé le fichier sur une disquette pour le donner à Alison Blackshaw[8].

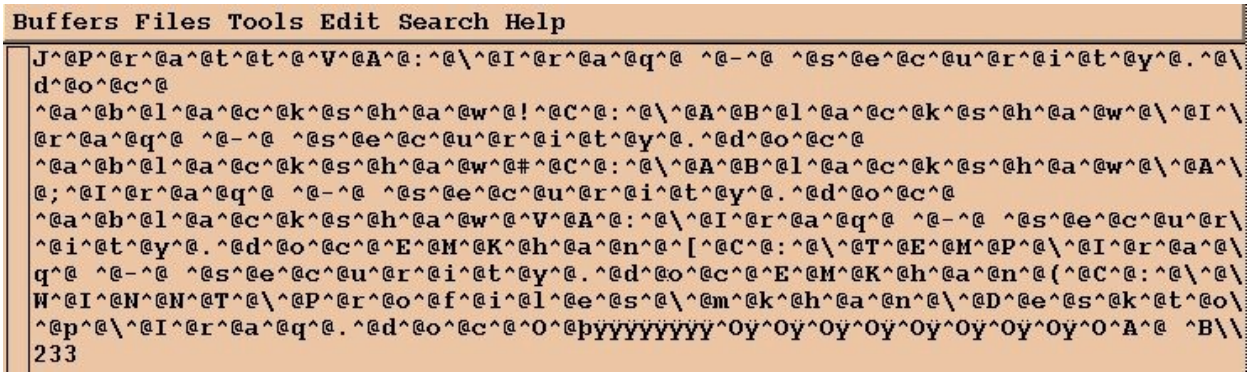

Figure 9 : visualisation d'une partie du document http://www.computerbytesman.com/privacy/blair.doc avec un éditeur de texte

La stabilité du contenu informationnel

Au-delà de la lisibilité, l'autre critère retenu est la stabilité du contenu informationnel. Nous venons de voir en effet que le maintien de la lisibilité sur le moyen et long terme d'une information peut imposer l'obligation de procéder à ces migrations de format qui transforment la structure profonde du fichier informatique (la suite de 0 et de 1 est modifiée). Ces migrations à l'inverse ne doivent pas affecter le contenu informationnel portée par le document, la donnée. La stabilité désigne ainsi « la nécessité de pouvoir garantir que les informations véhiculées par le document restent les mêmes depuis l'origine, qu'aucune n'est omise ou rajoutée. Le contenu informationnel s'entend de l'ensemble des informations et notamment de sa mise en forme ». Ceci est un point capital et extrêmement capital, qui requiert toutes les compétences archivistiques du professionnel qui va ainsi opérer et une collaboration fructueuse avec celui qui a produit l'information. Les questions suivantes se posent : le formalisme de tel document doit-il être impérativement conservé ? si l'affichage d'un document résulte d'un contenu intégré dans un fonds de page, peut-on conserver à part le fond de page et les contenus ? un document est-il composé d'une superposition d'informations qui, si on opère des migrations, risque d'en faire disparaître certaines ? les formules mathématiques présentes dans tel document vont-elles résister à telle migration ? On le voit, cela implique d'être très attentif concernant les formats que l'on va choisir ou vers lesquels on va migrer les documents que l'on a reçus. La stabilité du contenu informationnel est donc une question extrêmement délicate.

La traçabilité des opérations

Le troisième critère de l'intégrité telle qu'entendue dans la recommandation du Forum des droits sur l'Internet est la traçabilité. On retrouve ce mot dès lors que l'on parle d'administration électronique ou plus généralement numérique. Il s'agit là encore d'une notion fondamentale. Qui dit traçabilité, dit

[8] Pour plus de détails, http://www.computerbytesman.com/privacy/blair.htm .

auditabilité du service d'archive. Le processus d'archivage doit pouvoir être contrôlé tant du point de vue fonctionnel et organisationnel, que du point de vue technique et de la sécurité.

Qui dit traçabilité, implique qu'en amont, les rôles et responsabilités des différentes parties en présence soient explicitement posés. Ainsi un contrat (convention de service) doit-il être impérativement passé entre celui qui transfère des documents et données à archiver et le service d'archives. La question des formats et des migrations doit évidemment être clairement posée : formats acceptés en entrée, conversions et migrations admises. Le service d'archives doit par exemple s'engager à ne pas procéder à de telles migrations sans tracer l'ensemble des opérations qui sont effectuées (tests, résultats des tests, quel opérateur s'en est occupé, réversibilité des migrations, conservation ou non des documents et données dans le format originel....).

C'est ainsi qu'une étude a été conduite sous l'égide de la direction centrale de la sécurité des systèmes d'information (DCSSI)[9] (avec la collaboration de la direction des Archives de France et de la direction générale de la modernisation de l'Etat) qui a établi une « Archivage électronique sécurisé. P2A Politique et pratiques d'archivage (sphère publique) » qui fixe les exigences tant organisationnelles que fonctionnelles, juridiques et techniques permettant à des informations juridiques ayant une valeur probante, de conserver celle-ci tout au long du processus d'archivage. A partir de cette politique, une grille d'audit a été élaborée permettant des audits du système.

La conservation des documents signés

Signature électronique et signature cryptographique

Parmi tous les documents numériques à conserver, ceux sur lesquels une signature a été apposée posent des problèmes spécifiques. Il est important de noter que la problématique de la conservation de documents signés doit être distinguée de celle, complètement différente, de l'utilisation de procédés de signature électronique comme outils pour garantir l'intégrité au cours de la conservation d'un document.

La notion de signature électronique est définie à l'article 1316-4 du Code civil : « la signature lorsqu'elle est électronique consiste en l'usage d'un procédé fiable d'identification garantissant son lien avec l'acte auquel elle s'attache ». Il s'agit d'une définition très générale qui, d'un point de vue technique, peut recouvrir quantité de formes différentes : une signature manuscrite représentée sous forme numérique, une signature d'un document numérique qui, elle aussi, sera numérique, etc. Il faut donc veiller à distinguer la signature électronique telle qu'elle est définie par le Code civil et la signature cryptographique (ou signature à clef publique) qui est une technique particulière qui remplit les fonctions de signature électronique.

Les techniques de signature cryptographique produisent en effet des données numériques qui sont calculées à partir de la suite binaire qui représente le document numérique et des données de création de signature (clef privée) propres au signataire. Une signature obtenue par un tel procédé satisfait donc deux des exigences de la signature électronique sécurisée, définies dans le décret du 30 mars 2001 : elle garantit un lien avec le document numérique auquel elle s'attache, et elle est propre à la personne qui détient les données de création de signature. La troisième exigence pour être une signature électronique sécurisée fixée par le décret du 30 mars 2001 impose également que les données de création de signature soient protégées physiquement au moyen d'un matériel sécurisé,

[9] http://www.ssi.gouv.fr/fr/confiance/archivage.html

comme une carte à puce.

```
-----BEGIN PGP SIGNATURE-----
Version: GnuPG v1.0.7 (GNU/Linux)

iD8DBQA87jDM8dLMyEl6F20RAi1uAJ91R5Ca6JBguqv+02851N6pIYj+tACgvYt5
ANtRHJueYuGmb1zqa7sHJr4=
=yarZ
-----END PGP SIGNATURE-----
```

Figure 10 : exemple de signature cryptographique

Conservation de la signature cryptographique

Le fait qu'une signature cryptographique soit une chaîne binaire qui semble aléatoire à qui ne dispose pas du procédé et des données de vérification de signature (voir Figure 10) implique que, contrairement au cas de la signature manuscrite, son existence n'apporte aucune information tant qu'elle n'a pas été vérifiée. En effet, n'importe quelle suite binaire choisie au hasard ressemble à une signature cryptographique valide. Supposer qu'une signature cryptographique est a priori valide est donc impossible. Lors de l'archivage d'un document signé avec un procédé cryptographique, il paraît donc indispensable de conserver la capacité de vérifier cette signature au cours du temps. Cependant, cette exigence se heurte à trois problèmes cruciaux :

- la procédure de vérification de signature est une fonction qui prend en entrée un document numérique, une signature, et les données de vérification d'une signature d'une personne (la clef publique), et qui répond à la question « cette signature est-elle la signature émise à partir de ce document numérique par la personne dont voici la clef publique ? ». Autrement dit, pour prouver que la signature a été apposée par une personne donnée, il faut également vérifier le lien entre la clef publique et l'identité de son propriétaire. Le problème est ici similaire à celui des opérations réalisées avec une carte bancaire : les procédés techniques utilisés garantissent qu'une opération a été effectuée au moyen d'une carte donnée, mais ils ne prouvent en rien que cette opération a été effectuée par un individu donné ; pour cela, il faut également prouver que cet individu est détenteur de la carte bancaire considérée, que cette dernière n'a pas été volée... Dans le cas de la signature, le lien entre la clef publique et l'identité du signataire est réalisé par un certificat de clef publique. Ainsi, tout utilisateur, avant de pouvoir apposer pour la première fois une signature, doit s'enregistrer auprès d'une autorité d'enregistrement qui va vérifier son identité et va demander l'émission d'un certificat à une autorité de certification, comme prévu par le décret du 30 mars 2001 et l'arrêté du 26 juillet 2004. En pratique, chacun peut visualiser les différents certificats le concernant et les certificats qu'il a acceptés dans son navigateur (fonction « view certificates » ou « certificate manager »). Ainsi, dans l'exemple de la figure 11, on trouve le champ « RSA Public Key » qui donne la clef publique de l'utilisateur, c'est-à-dire les données qui permettent de vérifier toute signature qu'il aura apposé, mais aussi le champ « Issuer » qui désigne l'autorité ayant émis le certificat (ici l'INRIA).

Figure 11 : exemple du début d'un certificat de clef publique

```
Certificate:
    Data:
        Version: 3 (0x2)
        Serial Number: 43 (0x2b)
        Signature Algorithm: md5WithRSAEncryption
        Issuer: C=FR, O=INRIA, CN=INRIA-Standard
        Validity
            Not Before: Jun 30 20:29:40 2006 GMT
            Not After : Jun 30 20:29:40 2007 GMT
        Subject: C=FR, O=INRIA, OU=Lorraine, CN=Helene Kirchner/emailAddress=Helene.Kirchner@loria.fr
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (2048 bit)
                Modulus (2048 bit):
                    00:c9:1d:72:22:d4:84:eb:e4:65:fb:96:5d:df:57:
                    f9:b7:4f:cf:a4:ef:6f:1d:28:76:39:1f:49:c5:f9:
                    05:44:16:eb:e6:e2:43:36:b9:0b:35:36:af:01:ab:
                    5f:2b:fe:a6:ab:24:da:8b:18:fe:ac:01:85:7d:e5:
                    99:3e:cc:d2:53:17:bf:fe:73:83:8b:64:23:db:b9:
                    fa:d6:90:d6:a2:45:fa:91:bb:a0:bb:5e:02:f7:86:
                    23:de:68:f4:69:d9:f8:51:98:5f:88:a2:97:6e:f8:
                    05:77:cc:60:41:ab:9a:aa:b1:a4:dd:10:d2:4c:52:
                    e3:9c:38:cf:20:a4:b1:b5:20:1d:2e:04:cf:25:60:
                    be:9b:cd:39:b0:98:d3:28:88:22:09:14:b4:6c:db:
                    89:67:35:01:96:60:ed:5b:2b:42:36:64:42:f5:71:
                    7f:27:a1:43:fe:10:f6:37:89:08:ac:de:62:a6:1b:
                    60:b4:92:c2:25:ac:52:e3:8f:29:4d:6a:39:f1:5e:
                    23:bf:62:f3:5a:09:2b:75:91:de:79:5d:13:2c:0f:
                    6f:9b:23:e4:12:44:6d:75:30:50:54:76:21:11:f7:
                    35:cb:5e:2f:6d:05:17:cb:fa:c3:ac:dd:10:f1:81:
                    ae:19:ca:8a:72:35:70:c7:9c:82:e6:42:93:77:67:
                    b9:61
                Exponent: 65537 (0x10001)
        X509v3 extensions:
            X509v3 Basic Constraints: critical
                CA:FALSE
            Netscape Cert Type:
                SSL Client, S/MIME, Object Signing
            X509v3 Key Usage: critical
                Digital Signature, Non Repudiation, Key Encipherment
            Netscape Comment:
                Certificate INRIA-Standard. For more information, see http://igc.national.inria.fr/INRIA-Standard/
            X509v3 Subject Key Identifier:
                5A:F4:E7:22:AF:2B:2B:78:31:42:AC:5D:89:B7:51:E9:DD:BA:D9:F7
            X509v3 Authority Key Identifier:
                keyid:39:6F:D3:67:29:1B:39:1E:C9:EF:75:A0:7C:10:A9:8D:3B:4B:71:4F
                DirName:/C=FR/O=INRIA/CN=INRIA
                serial:02
```

Le certificat proprement dit est une signature de la clef publique de l'utilisateur par cette autorité. Pour pouvoir lui accorder la moindre confiance, il faut donc le vérifier, ce qui nécessite l'obtention d'un certificat pour cette autorité. On doit ainsi reconstituer une chaîne de certificats en remontant jusqu'à un certificat racine en lequel on peut avoir confiance. Par exemple, les certificats racines du Minefi sont mis à disposition sur le serveur http://www.finances.gouv.fr/dematerialisation_icp/VERIFICATION.htm et peuvent être vérifiés au moyen d'une empreinte publiée au Journal Officiel du 23 décembre 2004 ou lue par un serveur téléphonique vocal. Une deuxième vérification est cependant nécessaire avant de faire le lien entre l'identité d'une personne et sa clef publique : il faut vérifier que cette clef n'a pas été compromise, qu'elle ne figure pas sur une liste de révocation de clefs, et effectuer la même opération pour les clefs de chacune des autorités ayant émis un des certificats de la chaîne. Ainsi, si l'on veut garder la possibilité de vérifier une signature cryptographique au cours du temps, il faut non seulement archiver le document numérique auquel elle se rattache et la clef publique du signataire, mais aussi l'ensemble de la chaîne de certificats indispensables pour reconstituer le lien entre l'identité du signataire et sa clef publique, et l'ensemble des listes de révocation en vigueur à l'instant où la signature et les divers certificats ont été émis pour chacune des autorités de certification impliquées.

- Il est important de garder à l'esprit que le signature cryptographique est un procédé qui reste relativement vulnérable dans la durée : la clef propre à chaque utilisateur peut en effet lui être subtilisée, mais le procédé cryptographique de signature peut également perdre sa fiabilité du fait d'évolutions techniques. Ainsi, la sécurité de la signature RSA, qui est la technique la plus utilisée, repose sur l'impossibilité pratique de factoriser un nombre entier formé par le produit de deux grands nombres premiers. Par exemple, on ne sait pas actuellement résoudre le problème suivant[10] : sachant que le nombre suivant de 212 chiffres décimaux n s'écrit comme un produit n=p*q, retrouver les deux facteurs p et q, pour n = 7403756347956171282804679609742957314259318888923128908493623263897276503402826627689199641962511784399589433050212758537011896809828673317327310893090055250511687706329907239638078671008609696253793465056379635 9 . Toutefois, la taille du plus grand nombre de ce type que l'on sait factoriser augmente au cours du temps grâce à l'augmentation de la puissance des ordinateurs et aux progrès mathématiques dans ce domaine. Ainsi, un nombre de 512 bits (155 chiffres décimaux) a été factorisé en 1999, alors que ceci semblait hors de portée dans les années 1980[11]. C'est ainsi que nos cartes bancaires utilisaient depuis leur conception des signatures RSA avec des nombres de 320 bits, et que leur sécurité, relativement raisonnable lors de la mise en service du système, a été mise à mal en 1999 (par l'attaque de Serge Humpich qui a montré qu'il était possible de contrefaire des cartes), ce qui a nécessité d'augmenter la taille des nombres utilisés. Il est donc clair que l'horodatage sécurisé de toute signature cryptographique (mais aussi des certificats par exemple) est indispensable : une signature émise avec un procédé fiable à un instant donné pourra avoir entièrement perdu sa sécurité 10 ou 20 ans après. Il faut anticiper la vérification de la signature en amont de sa conservation.
- Le dernier point important est que la signature cryptographique perd entièrement sa validité dès que l'on change le format, la version du format ou le codage du document auquel elle se rattache. En effet, différents documents qui ont pourtant le même contenu perceptif, comme ceux que nous avons vus auparavant, ont des signatures cryptographiques totalement différentes dans la mesure où celles-ci ne dépendent que de la suite binaire représentant le document. Garder la possibilité de vérifier une signature cryptographique au cours du temps interdit donc toute migration de format, opération pourtant nécessaire pour assurer la lisibilité du document.

La réponse de l'archiviste

Il y a cette difficulté intrinsèque mais qui est la même que celle que nous évoquions tout à l'heure : on ne peut pas figer les choses dans le temps, dès lors qu'on prétend vouloir pérenniser une information numérique. Or, comme l'a souligné Jean-François Blanchette, la cryptographie est une solution à un problème qui se pose dans l'espace, mais dans le temps.

Les migrations de format que nous avons évoquées ci-dessus invalident nécessairement les signatures originelles des documents migrés. Il est dès lors impossible dès lors qu'on se situe dans le temps

[10] Il s'agit d'un des défis lancés par la société RSA Inc et disponibles sur http://www.rsa.com/rsalabs/node.asp?id=2093 é

[11] Pour des détails sur l'évolution au cours du temps de la taille du plus grand nombre factorisé, on peut se reporter à : François Morain, *La factorisation d'entiers*, dossier hors série de Pour la Science « L'art du secret », pages 62-64, juillet-octobre 2002, et à la page web de Paul Zimmermann, http://www.loria.fr/~zimmerma/records/factor.html

d'imaginer un système d'archivage pouvant autant que nécessaire (2, 5, 10, 30, 100 ans...) garantir la signature originelle des documents reçus pour archivage..

Une telle exigence serait non seulement impossible techniquement (pour les documents migrés), extraordinairement complexe et coûteuse (pour les documents non migrés, si on voulait conserver toute l'infrastructure logicielle, les certificats, les jetons nécessaires à la signature originelle du document, ou même si on s'engageait dans des procédures de re-signature) et fonctionnellement absurde : demande-t-on aujourd'hui à un archiviste de vérifier les signatures manuscrites qui figurent sur les documents qu'il reçoit ? ce qu'on demande à l'archiviste, c'est de garantir le processus d'archivage y compris paradoxalement pour les documents qui seraient des faux : l'archiviste doit pouvoir prouver qu'il a parfaitement conservé des faux !

Par conséquent, la réponse à la question de la conservation des documents originellement signés, est une réponse d'ordre organisationnelle : c'est aux services qui ont produit ou reçu les documents signés de préparer la preuve de ce qui pourra être du contentieux plus tard : c'est à eux par conséquent d'introduire, avant leur transfert dans un service d'archives, une procédure de vérification de documents signés et de porter les résultats de cette vérification dans les métadonnées qui accompagnent le document. Cette vérification doit dans l'idéal être réalisée rapidement après la production du document, afin de pallier l'obsolescence du certificat. Le service d'archives va alors archiver le document avec ses fichiers de signature, mais sans la vérifier. Par contre, c'est le processus d'archivage qui devra pouvoir être vérifié : empreintes à générer sur les fichiers transférés et contrôles d'intégrité à effectuer à l'entrée dans le système puis de manière régulière et systématiquement lors d'une demande de communication/restitution ; signature électronique éventuellement pour des messages engageant la responsabilité des archivistes et services (acceptations de versement, ou encore autorisations de destructions..).

L'autre type de réponse, juridique, a été donnée, on l'a dit plus haut, avec les décrets du 13 août 2005 relatifs aux actes authentiques électroniques. les migrations successives nécessaires pour assurer la lisibilité des documents ne lui enlèvent pas son statut d'origine. Cette précision apporte une sécurité juridique au dispositif de conservation.

Le processus d'archivage

L'objectif qui tend à assurer l'intégrité d'un document ne peut être obtenu, nous venons de le voir, qu'en se reposant sur un processus d'archivage. La mise en oeuvre de ce processus permettra aux acteurs de donner leur confiance dans la façon dont ces documents ont été pris en charge et conservés.

Le processus est organisé autour de quelques grandes actions : organiser, capturer, conserver, communiquer. Différents référentiels, guides de bonnes pratiques, outils, organisations permettront de répondre à ces différentes problématiques. Quant à l'organisation à mettre en place (transfert vers un service d'archive en interne, externalisation vers un tiers-archiveur, mutualisation notamment du stockage entre plusieurs organismes…), tout dépendra des enjeux et du contexte.

La politique d'archivage élaborée par la DCSSI, évoquée plus haut, pose le cadre et permet à partir de cet outil d'élaborer une « déclaration des pratiques d'archivage » qui explicite les moyens mis en

oeuvre pour répondre aux objectifs posés par la politique d'archivage. Elle permet notamment comme précisé plus haut, de fixer les responsabilités des différentes parties prenantes au processus et ainsi de pouvoir, par grand type de documents ou de données transférés, spécifier des contrats.

Concernant le transfert des documents et données depuis un service dit service versant vers un service d'archives, la DAF et la DGME ont élaboré en mars 2006, un standard d'échange de données pour l'archivage visant à faciliter l'interopérabilité des échanges entre un service d'archives et ses différents partenaires[12]. Concernant le transfert notamment, le standard prévoit l'échange d'un certain nombre de messages mais précise également comment les données doivent être transférées et quel doit être le contenu de ce transfert : l'ancien bordereau de versement papier est ainsi modélisé et toutes les informations (métadonnées de gestion, descriptives, techniques) sont précisées qui doivent accompagner les données durant le transfert. Ainsi le standard permet-il à partir des systèmes d'information métier d'origine d'effectuer des exports au format du standard, c'est-à-dire permet de spécifier comment sont fabriquées et ce que contiennent les enveloppes XML transférées (données et métadonnées).

Le transfert va s'effectuer suivant les termes du contrat liant le service d'archives au service dit service versant, en précisant un certain nombre de points : modalités du transfert (réseau ou pas ? quel protocole ?), fréquence et périodicité des versements ? format d'exportation des données ? comment celles-ci vont s'ordonner ? quelles informations vont comporter les messages ? ceux-ci seront-ils chiffrés ? dans quel ordre vont se trouver les informations ? comment reconnaît-on cet ordre ? délais de conservation des documents, clauses particulières de confidentialité, délais de communicabilité, modalités de restitution…

Les enveloppes, une fois transférées, seront vérifiées et contrôlées et une fois le versement accepté par le service d'archives, les données seront d'une part écrites dans l'espace de stockage et une partie des métadonnées récupérées par le système de gestion documentaire.

En ce qui concerne la conservation proprement dite, elle englobe plusieurs types de fonctions. Au-delà des principes généraux de sécurité qui s'appliquent pour tout système d'information, un principe fondamental concernant la conservation de données numériques pour leur archivage est d'accroître la sécurité grâce à la redondance et à la réplication : redondance des données, des applicatifs mais également des accès, et surtout réplication des données sur deux sites distants, ceci sans préjudice des politiques de sauvegarde traditionnelles afférentes à tout système d'information. Quant aux questions sur les types d'infrastructure de stockage, le choix de types de supports, tout dépendra du contexte. Ce qu'il importe avant tout, c'est de choisir un système robuste, évolutif, permettant des contrôles d'intégrité des données, et qui permette l'abstraction de la plate-forme matérielle (pouvoir évoluer sans impact sur l'organisation logique des archives). Tout système doit faire l'objet d'une surveillance régulière, qu'on peut plus ou moins automatiser suivant les volumes et les technologies choisies. Par ailleurs, s'il y a de forts enjeux juridiques, on pourra choisir des technologies dites Worm qui permettent de détecter toute modification qui aurait été effectuée sur un support après sa gravure ou son écriture initiale.

Dernière fonction : la communication, la diffusion, à partir des métadonnées intégrées dans le système de gestion documentaire du système. La communication des documents et données archivés

[12]https://www.ateliers.modernisation.gouv.fr/ministeres/projets_adele/a103_archivage_elect/public/standard_d_echange_d/folder_contents

à une communauté plus large que celle qui l'a produite, nécessite généralement d'apporter des éléments d'information complémentaires par rapport à ce qui a été transmis au titre des métadonnées accompagnant les données. Les modalités pratiques de la communication (en ligne, différé) sont évidemment à préciser suivant le type de public et le type de données.

Enfin un processus d'archivage comporte un volant « administration » du système (comprend notamment un journal des événements indispensable à la traçabilité telle qu'elle a été évoquée plus haut), ainsi que des volets planification et pérennisation (veille technologique, planification des migrations de supports, des formats).

Conclusion

Les réponses univoques à un problème sont toujours insatisfaisantes. Si on pense qu'une technologie donnée va résoudre, à elle seule, le problème de l'intégrité et de la conservation de l'information numérique, il est probable qu'on va se tromper. De la même manière qu'une lecture uniquement juridique de la question risque d'amener à des impasses. La réponse est par conséquent une approche pluridisciplinaire qui permettra de faire travailler ensemble informaticiens, juristes, archivistes et ainsi de pouvoir construire une réponse à la fois organisationnelle, fonctionnelle, technique et juridique.