



# Les formats de fichiers PDF

## Structure – Génération – Utilisation

Références GF/GFFC/120124

Septembre 2012

**Gérald FAVRE**

**FAVRE Consulting SARL – Siège Social** : 103 Avenue Aristide BRIAND 93190 LIVRY GARGAN

Tél : 01 43 30 12 12 Télécopie : 01 43 30 99 00 Portable : 06 13 08 07 27

RCS Paris B 451 229 686 000 30 Code APE 741G Code NAF 7022 Z

Numéro de TVA intracommunautaire : FR41451229686

Application de l'article 259 B du Code Général des Impôts

e-mail : [gerald.favre@favre-consulting.net](mailto:gerald.favre@favre-consulting.net)



**Résumé :**

Ce document offre un ensemble d'informations de base, compréhensible par les non-spécialistes. Son objectif est de permettre de fonder une bonne utilisation des fichiers au format PDF et de corriger certaines erreurs d'appréciations courantes. Il est destiné principalement aux créateurs et aux gestionnaires de documents électroniques.

**Abstract :**

This document offers background information to non-specialists, in order to asses the proper use of PDF format files and to correct some frequent misinterpretations. It is aimed at electronic documents creators and managers.



## SOMMAIRE

<b>1. INTRODUCTION .....</b>	<b>5</b>
1.1. Présentation .....	5
1.2. Sigles .....	5
1.3. Définitions .....	6
<b>2. ORIGINE DU FORMAT .....</b>	<b>8</b>
2.1. Historique du format PDF .....	8
2.2. Evolution du format PDF dans ses différentes versions .....	8
2.3. Les formats concurrents : HTML et XPS.....	9
<b>3. LA STRUCTURE DES FICHIERS PDF .....</b>	<b>10</b>
3.1. Aperçu du format PDF propriétaire .....	10
3.1.1. La structure du fichier .....	10
3.1.2. La structure du corps du document .....	12
3.1.3. Autres contenus du fichier.....	12
3.1.4. Les formats de police de caractère .....	12
3.2. Schéma de création d'un fichier PDF .....	13
3.2.1. Création à partir d'un outil logiciel .....	13
3.2.2. Création à partir d'un matériel de numérisation.....	14
3.3. Les formats de fichiers PDF à vocation normative .....	14
3.3.1. PDF/A.....	15
3.3.2. PDF/X.....	15
3.3.3. PDF/E .....	17
3.3.4. PDF/UA.....	17



<b>4. OUTILS LOGICIELS TRAITANT LE FORMAT PDF .....</b>	<b>18</b>
4.1. Outils logiciels du marché.....	18
4.2. Outils logiciels du monde libre .....	18
4.2.1. Outils logiciels de création de fichiers PDF .....	18
4.2.2. Outils logiciels de création de fichiers à partir du format PDF.....	19
4.3. Outils logiciels de validation des formats .....	19
<b>5. L'APPORT DE LA VERSION 9 D'ADOBE® ACROBAT® .....</b>	<b>20</b>
<b>6. LA SIGNATURE ELECTRONIQUE .....</b>	<b>21</b>
6.1. La signature électronique définie par les normes internationales .....	21
6.1.1. Aspects légaux de la signature électronique.....	21
6.1.2. Propriétés de la signature électronique.....	22
6.2. Principe des mécanismes de signature .....	22
6.2.1. Les signatures non reconnues légalement .....	22
6.2.2. Les signatures légales.....	23
6.3. Transmission sécurisée de documents électroniques .....	23
<b>7. RECOMMANDATIONS .....</b>	<b>25</b>
7.1. Avenir du format PDF.....	25
7.2. Sécurisation du format PDF .....	25
7.3. Interprétation des fichiers PDF .....	25
7.4. Utilisation des différents formats PDF.....	26
7.5. Taille variable des fichiers PDF créés.....	26
7.6. Lenteur de l'impression de certains documents PDF.....	27
7.7. Archivage électronique et format PDF.....	27
<b>8. REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>28</b>



## **1. INTRODUCTION**

### **1.1.PRESENTATION**

Le format PDF (en réalité, nous verrons qu'il y en a plusieurs) facilite le partage et la sécurisation de documents sur support électronique, sous réserve de connaître et de disposer des bons outils de gestion (création, production, mise à jour) et de transfert.

Ce document a été produit avec l'objectif de fournir aux utilisateurs des systèmes de gestion de l'information documentaire et aux créateurs de documents les informations nécessaires leur permettant de choisir en toute connaissance de cause quel format et quel outil utiliser pour un usage particulier.

Les informations contenues dans ce document ont été mises à jour lors de sa rédaction. Elles prennent en compte la structure de fichier PDF 1.7 et son évolution récente, utilisée par les dernières versions des outils d'Adobe® Acrobat® et des outils qui lui sont associés. Elles devront être complétées à mesure qu'apparaîtront de nouvelles versions, de nouveaux outils ou de nouveaux formats.

### **1.2.SIGLES**

AIIM : Association for Information and Image Management

ANSI : American National Standards Institute

ASCII : American Standard Code for Information Interchange

DCSSI : Direction Centrale de la Sécurité des Systèmes d'Information

DGME : Direction Générale de la Modernisation de l'Etat

ECMA International : European association for standardizing information and communication systems

EDI : Echange de données informatisé

GEIDE : Gestion électronique des informations et documents entreprise

HTML : HyperText Markup Language

ISO : International Standard Organization

PDF : Portable Document Format

PS : PostScript



RCO : Reconnaissance optique de caractère

RVB : Rouge, Vert, Bleu

SGML : Standard Generalized Markup Language

SIG : Système d'Information Géographique

XML : eXtensible Markup Language

XPS : XML Paper Specification

### 1.3.DEFINITIONS

**Le format PDF** existe sous deux structures :

- La structure texte qui est un fichier généré par une application bureautique. Il est constitué de caractères. C'est celui qui est couramment désigné par « PDF interprétable ».
- La structure image qui est un fichier issu d'un scanner ou d'outils logiciels de création d'image (type Photoshop). Il ne peut pas être exporté vers des outils bureautiques et son contenu textuel est inaccessible en tant que tel.

**Le format PDF 1.?** (« ? » peut prendre les valeurs 1, 2, 3, ... jusqu'à 7 aujourd'hui) est le format propriétaire des fichiers PDF que la société Adobe a spécifié pour ses outils. Les autres formats de fichiers : PDF/A, PDF/X, PDF/E, PDF/UA (voir chapitre 3) sont des formats standards développés par une communauté d'éditeurs de logiciels et dont les spécifications sont disponibles, voire normées. Ces formats s'appuient sur le format PDF propriétaire d'Adobe®.

**Le format XPS** est un format de fichier développé par Microsoft en concurrence au format PDF et en cours de standardisation auprès de l'ECMA. Ce format de fichier a été intégré en 2007 dans la version XP du système Windows®.

**La signature électronique** est un mécanisme qui permet d'authentifier l'auteur d'un document électronique et de garantir son intégrité. L'analogie avec une signature manuscrite est présentée dans l'article 1316-4 du Code Civil.

**Le format PostScript** inclut un langage vectoriel utilisé pour positionner les informations relatives à un document en vue de leur utilisation par les outils d'impression.

**Le langage SGML**, langage normalisé de balisage généralisé, adopté comme standard en 1986 (ISO 8879), a été la première tentative systématique de créer de réels documents électroniques, c'est-à-dire des documents qui n'étaient plus des documents papier repris sous forme électronique. La principale idée sous-jacente



était de séparer le *contenu* (logique) d'un document de sa *forme* (matérielle/imprimée). L'intention finale était toujours *principalement* de produire des documents imprimés, en économisant les ressources informatiques et en limitant les risques d'erreur — un unique document (logique) étant transformé automatiquement en différents formats imprimés (par exemple des documents de maintenance pour des avions produits en différents formats selon les normes de présentation de différentes compagnies aériennes). La première expérience de terrain en France fut la documentation de maintenance des sous-marins nucléaires. SGML a constitué une innovation, mais il était si complexe que sa manipulation s'est trouvée restreinte aux spécialistes (rédacteurs techniques dans l'industrie ou spécialistes textuels dans les sciences humaines).

SGML est un langage très riche, mais très complexe dans sa mise en œuvre et dans son utilisation.

**Le langage HTML**, langage de balisage hypertexte conçu vers 1990, est devenu une réalité à partir de 1992, comme le Web à la même date. Tous deux ont popularisé les documents électroniques hypertextuels sur une immense échelle. HTML utilise une bibliothèque de balises très générale, mais figée et réduite, ce qui *in fine* n'était plus adapté aux contraintes nouvelles du Web imposées par l'évolution des outils. Ces évolutions demandaient un langage disposant d'une sémantique de plus en plus riche. C'est pour cela que HTML a été déplacé sur le Web par XML, qui permet un plus grand niveau de spécialisation par domaines.

**Le langage XML**, langage de balisage extensible, est le langage destiné à succéder à HTML sur le Web. C'est un langage de balisage (markup), c'est-à-dire qui présente de l'information encadrée par des balises. Mais contrairement à HTML qui présente un jeu limité de balises orientées présentation (titre, paragraphe, image, lien hypertexte, etc.), XML est un métalangage qui permet d'inventer à volonté de nouvelles balises pour isoler toutes les informations élémentaires (titre d'ouvrage, prix d'article, numéro de sécurité sociale, référence de pièce...) ou les agrégats d'informations élémentaires que peut contenir une page Web.

XML est un sous-ensemble au sens strict de SGML, dont il ne retient pas les aspects trop ciblés sur certains besoins. En cela il représente un "profil d'application" de la norme SGML.



## 2. ORIGINE DU FORMAT

### 2.1. HISTORIQUE DU FORMAT PDF

Le format PDF a été créé au début des années 90 par la société Adobe, en même temps que la famille de produits Adobe® Acrobat®. A cette date le langage de description de page PostScript® s'est rapidement imposé au niveau mondial comme le langage standard de présentation et d'impression des documents.

Le format PDF a été initialement créé pour permettre la lecture d'un document identique à celle d'un support papier. En effet les formats utilisés pour la présentation à l'écran provenaient de ceux utilisés pour l'impression : principalement le format PostScript® qui modifie les couleurs, les schémas (diminution de la taille de 3 %) et la présentation par rapport à l'original (voir § 3.2).

Le format PDF permet de respecter :

- les polices initiales de caractères, sous réserve de leur disponibilité dans la version utilisée de l'outil logiciel de création,
- la présentation et les schémas ou images contenues,
- la mise en forme du document.

Sur la base du respect de l'original qui reste la ligne directrice des créateurs du format PDF, ce dernier a évolué depuis ses origines vers une utilisation plus large que la présentation à l'écran ou l'impression : l'échange de documents par voie électronique et l'archivage des documents électroniques sur des périodes longues en minimisant les contraintes des outils de présentation.

Depuis maintenant une dizaine d'années, devant l'évolution exponentielle de l'utilisation d'Internet pour les échanges, le format PDF s'est établi comme un standard *de facto* pour les transferts de documents sur support électronique.

### 2.2. EVOLUTION DU FORMAT PDF DANS SES DIFFERENTES VERSIONS

Le format de fichier PDF a évolué depuis sa création en 1993, en association avec les outils développés par Adobe autour de lui. Le tableau ci-dessous donne la correspondance entre les versions PDF et celles d'Acrobat®.





Année de mise sur le marché	Version du format de fichier	Version des outils d'Adobe®
1993	Format PDF 1.0	Acrobat® 1.0
1994	Format PDF 1.1	Acrobat® 2.0
1996	Format PDF 1.2	Acrobat® 3.0
1999	Format PDF 1.3	Acrobat® 4.0
2001	Format PDF 1.4	Acrobat® 5.0
2003	Format PDF 1.5	Acrobat® 6.0
2005	Format PDF 1.6	Acrobat® 7.0
2006	Format PDF 1.7	Acrobat® 8.0
2008	Format PDF 1.7	Acrobat® 9.0

### 2.3.LES FORMATS CONCURRENTS : HTML ET XPS

Ce paragraphe est donné pour information car les enjeux d'un format unique au niveau mondial sont importants industriellement : d'autres formats de fichiers sont actuellement en concurrence aux Etats-Unis (moins connus en France) et risquent de prendre partiellement ou entièrement le marché couvert par Adobe et son format PDF.

Dans le domaine de la gestion électronique des documents (fonctions de transfert, d'affichage et de stockage à moyen terme) deux formats de fichiers se font concurrence : le format PDF et le format HTML. L'avantage de PDF, par rapport à HTML, tient au fait que la restitution du document est fidèle à l'original et qu'elle ne dépend ni de la configuration de la machine sur laquelle s'effectue la consultation, ni de la version du logiciel. Ce n'est pas le cas en HTML où les différences entre les navigateurs et dans le paramétrage se traduisent par des variations sur l'écran ou lors de l'impression.

Microsoft® a développé récemment pour ses outils le format XPS, qui est une déclinaison du langage XML dans l'environnement de MS Office. XPS se veut être un standard destiné à rivaliser avec PDF et a été déposé auprès de ECMA International, avec l'objectif de devenir un standard et postérieurement une norme.



## 3. LA STRUCTURE DES FICHIERS PDF

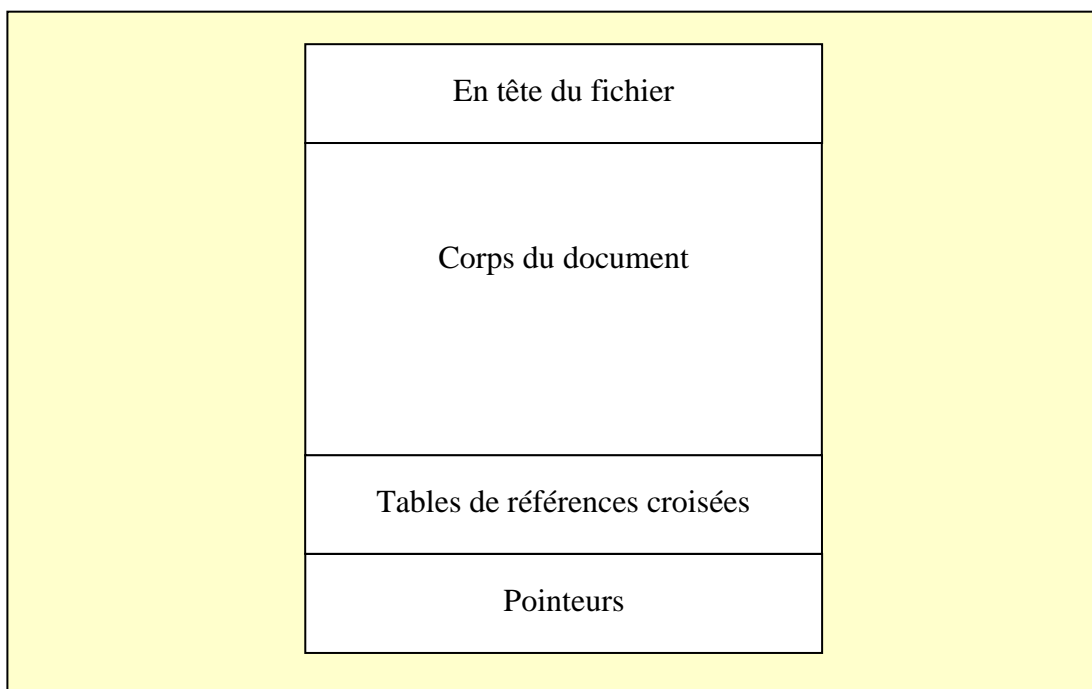
### 3.1. APERÇU DU FORMAT PDF PROPRIÉTAIRE

#### 3.1.1. LA STRUCTURE DU FICHIER

La description qui est donnée dans ce paragraphe est un aperçu de la structure d'un fichier PDF et ne prétend en aucun cas en être une description exhaustive<sup>1</sup>.

Un fichier PDF est structuré en 4 parties :

- L'en-tête du fichier qui indique, entre autres, la version de sa structure, la nature du corps du texte, etc.
- Le corps du texte qui est le contenu du document (texte, multimédia).
- Les tables de références qui sont destinées à faciliter l'accès à une page, un mot, un graphique, etc.
- Les pointeurs qui permettent d'accélérer les accès aux tables de références et à des objets particuliers contenus dans le corps du document.

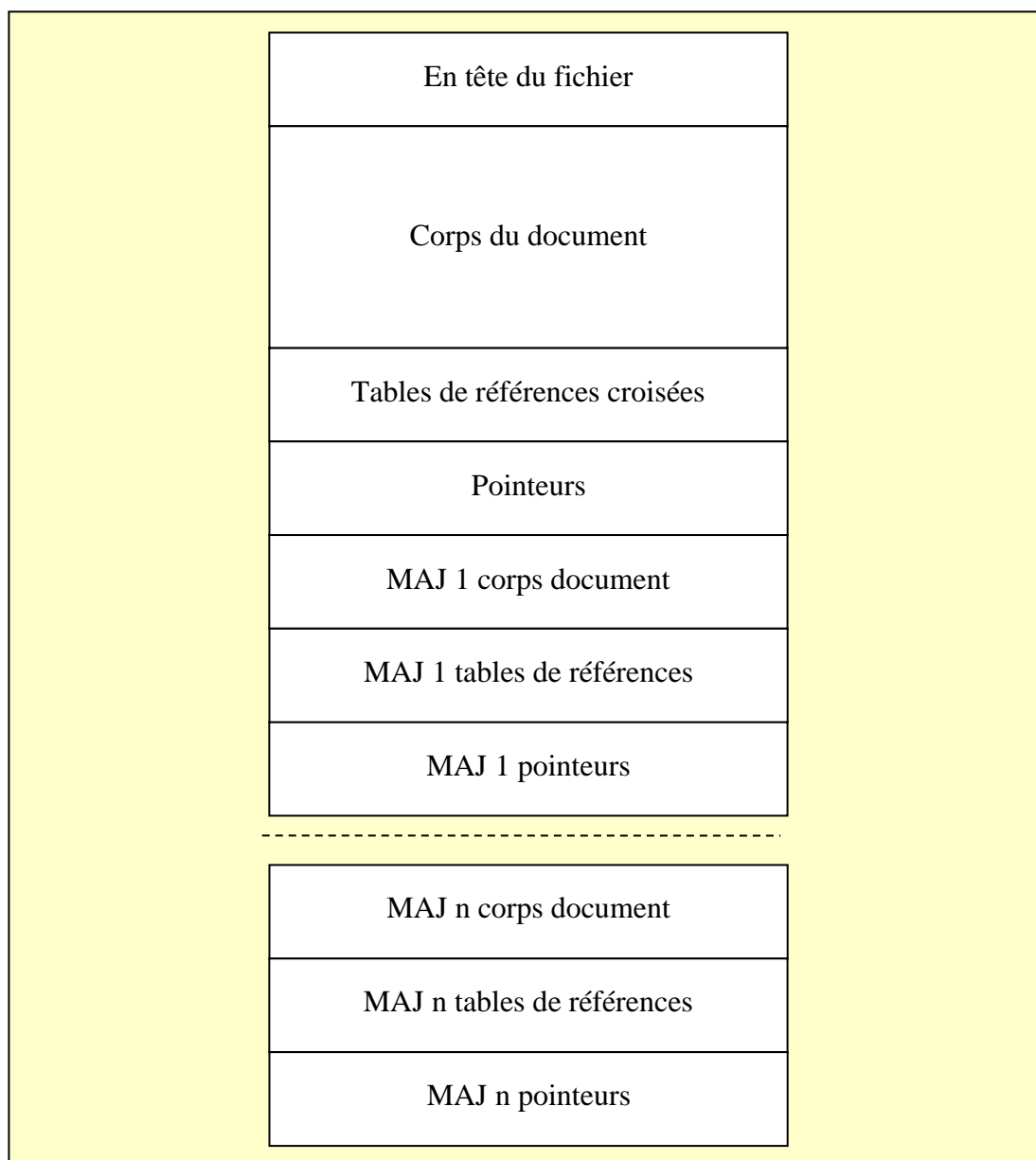


---

<sup>1</sup> Pour obtenir une description exhaustive de ce format, consulter le manuel de référence des informaticiens pour créer des fichiers PDF [voir référence bibliographique N°5].



Cette structure de fichier est stable pour toute utilisation (lecture, affichage, échanges, etc.), mais elle doit évoluer dans le temps pour être en adéquation avec les évolutions des structures de fichiers imposées par les outils de traitement et de gestion (création, affichage, etc.). Afin de parvenir à une structure évolutive qui suive les changements de versions générés par Adobe<sup>®</sup>, sans recréation du fichier de base, la technique utilisée est celle des ajouts de compléments par approches successives des nouvelles versions, ce qui conduit au schéma suivant :



**Remarque :** l'utilisation d'une structure incrémentale de fichier qui suit les versions successives des logiciels se rencontre dans la plupart des outils informatiques et bureautiques. Il est donc important, pour pouvoir lire intégralement des fichiers créés par des versions antérieures d'un outil, de suivre la séquence complète des mises à



jour (MAJ) de l'outil de création ou d'exploitation, ou bien d'acquérir à nouveau la version complète qui contient l'ensemble des modules du logiciel.

### **3.1.2. LA STRUCTURE DU CORPS DU DOCUMENT**

Le corps du document est lui aussi très structuré. Nous présentons dans ce paragraphe quelques aspects de cette structure pour des documents textuels. Les schémas, les contenu audio et vidéo ont eux aussi leur structure propre [voir réf. 5].

Les principales composantes de la structure des documents textuels sont :

#### 1/ Pour le caractère

Le caractère lui-même au format du standard ASCII et les attributs du caractère : la police, la taille, la couleur, le souligné, le gras, etc.

#### 2/ Pour le mot

Les attributs généraux donnés à la représentation du mot : l'espacement entre caractères, l'espacement horizontal entre les mots, l'espacement vertical, etc.

#### 3/ Pour la ligne

Les attributs du document qui correspondent à la feuille de style, mais aussi à des particularités comme le retour chariot (fin de ligne), la tabulation, etc.

### **3.1.3. AUTRES CONTENUS DU FICHIER**

Un fichier PDF peut aussi être compressé ou disposer d'une signature électronique. Ces informations sont identifiées dans le fichier afin d'informer l'outil de restitution des fonctions à appeler pour que cette restitution soit exacte et non entachée d'erreurs ou d'aberrations.

### **3.1.4. LES FORMATS DE POLICE DE CARACTERE**

Pour pouvoir être compatible avec la majorité des outils de gestion des documents (impression, création) le format PDF dispose en standard de 14 polices : Times, Courier, Symbol, Helvetica et Zap Dingbats avec leurs variantes italiques et grasses quand elles existent.

D'autres polices de caractères peuvent être intégrées dans les outils de création et d'affichage des fichiers PDF.

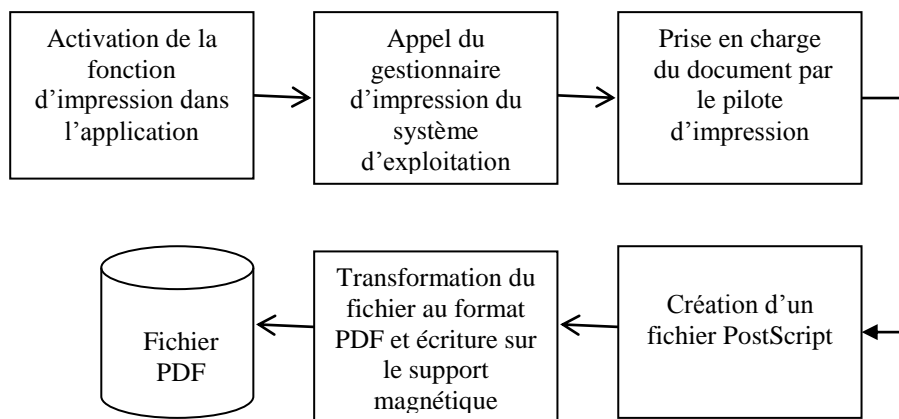


## 3.2. SCHEMA DE CREATION D'UN FICHIER PDF

### 3.2.1. CREATION A PARTIR D'UN OUTIL LOGICIEL

La fonction de création d'un fichier au format PDF consiste à imprimer un document à partir d'un logiciel de traitement de texte ou d'une application informatique, le résultat de cette impression étant un document semblable à l'original mais dans un format différent.

Le processus de création d'un document PDF à partir d'un outil logiciel de bureautique, quel que soit le système d'exploitation de la station de travail, est présenté dans le schéma suivant :



L'application qui a reçu l'ordre d'imprimer (par exemple Word®) active sa fonction d'impression afin de se mettre en liaison avec le gestionnaire d'impression du système d'exploitation qui gère les unités d'impression, les paramètres, les matériels associés, les priorités, les interruptions.

Le pilote (*driver* en anglais) prend à sa charge le document à imprimer pour le traduire, à partir du fichier d'origine créé par l'application, dans le format attendu par l'imprimante. Ce format d'impression est le format de fichier PostScript (extension .PS) qui permet de transmettre à l'imprimante les informations utiles de contenant et de contenu pour l'impression.

Cette même opération est utilisée pour la présentation sur les écrans : dès l'origine de l'apparition des écrans d'ordinateurs reliés à des unités centrales, le « masque d'écran » (MAP en anglais technique) a été considéré comme un fichier possédant des attributs spécifiques (positionnement, couleur, etc.) et non comme le transfert d'informations vers un matériel possédant ses propres règles de contrôle de fiabilité, d'écriture et d'affichage.



Avec la finalité d'obtenir un fichier affichable, le fichier PS est transformé en fichier PDF en y intégrant des attributs qui permettront l'utilisation de fonctions complémentaires comme l'annotation, la visualisation en 3D, etc.

L'utilisateur d'une application bureautique ne voit pas ce processus car il est généralement réalisé en mémoire vive, donc avec des traitements relativement rapides.

### **3.2.2. CREATION A PARTIR D'UN MATERIEL DE NUMERISATION**

Le fichier PDF issu d'un matériel de numérisation est créé par un logiciel spécifique intégré au matériel. Le matériel en soi ne produit aucun fichier sans logiciel.

Deux résultats différents sont produits selon le matériel utilisé : un fichier « image » à partir d'un scanner de bureau simple ou un fichier « texte » à partir d'une chaîne de numérisation complète. Une telle chaîne inclut l'ensemble des outils matériels et logiciels qui créent des fichiers PDF interprétables : après numérisation de l'image du document, un logiciel de reconnaissance optique de caractères (ROC) reconstitue les caractères et crée un fichier PDF dont le contenu textuel peut être exploré par un outil de traitement automatique de la langue.

## **3.3. LES FORMATS DE FICHIERS PDF A VOCATION NORMATIVE**

Il existe actuellement<sup>2</sup> quatre<sup>3</sup> grands types de formats PDF disponibles pour les éditeurs, qui ont été développés par un consortium de sociétés (entre autres l'AIIM) dont font partie Adobe, IBM, HP, etc. Ces formats sont normés ou en cours de normalisation et ne sont pas « propriétaires » de Adobe®. Ce sont : PDF/A, PDF/X, PDF/E, PDF/UA. Ils peuvent être utilisés par tout industriel ou particulier car ils sont libres de tout droit d'utilisation. Quant au PDF/X, il se décline en 4 sous catégories : X-1, X-2, X-3 et X-4 ; un X-5 est en cours d'étude.

La base technique qui a été prise pour la création et le développement de ces formats est celui du format de fichier PDF propriétaire qu'Adobe n'a que partiellement dévoilé pour garder ses droits et ses prérogatives sur ses outils.

---

<sup>2</sup> Les informations contenues dans ce chapitre ont pour origine principale celles fournies par la société Adobe®.

<sup>3</sup> D'autres formats de fichiers comme le PDF/H ou le PDF/IT ont été créés par des industriels pour des besoins propres et ne sont pas destinés à être normés ; ils ne seront donc pas traités ici.



### 3.3.1. PDF/A

Ce format fait partie de la norme ISO 19005-1 publiée fin 2005 pour l'archivage des documents électroniques. Il définit techniquement tous les éléments nécessaires pour éviter toute dépendance des fichiers vis-à-vis du matériel et des logiciels. Cela permet de respecter la norme ISO 15489<sup>4</sup> pour l'archivage à long terme des documents électroniques qui peuvent subir des évolutions de supports, d'outils de lecture, de recopie, etc. En effet, l'application des règles de la norme ISO 19005-1 assure la représentation fidèle du document PDF/A quel que soit l'environnement de lecture du fichier.

Tout éditeur de logiciel peut disposer, à des fins de développement d'outils, à la structure de ce format fichier PDF/A, ce qui n'est pas possible pour le format PDF 1.x qui est propriétaire et appartient à la société Adobe.

Ce format PDF/A est déjà accepté par certains organismes étatiques français :

- PIL@E : projet de mise en œuvre des concepts piliers de l'archivage électronique initialisé en 2005 par la Direction des Archives de France (DAF) ;
- DGME : elle recommande l'utilisation du PDF/A dans son Référentiel Général d'Interopérabilité (RGI) ;
- DGI : la Direction Générale des Impôts accepte la conservation au format PDF/A de la copie d'une facture envoyée à un client (cf. bulletin officiel des impôts du 11 janvier 2007).

Les éditeurs de solutions d'archivage adaptent leurs solutions au PDF/A pour être prêts à répondre aux demandes de leurs clients.

### 3.3.2. PDF/X

Ce format est principalement utilisé dans le domaine de l'édition pour les transferts de documents entre les créateurs et les imprimeurs. Il fait partie des recommandations ISO 15930<sup>5</sup>.

---

<sup>4</sup> La normalisation des principes et des procédures de "Records management", décrits dans la norme ISO 15 489, garantit que tous les documents d'archives bénéficient de l'attention et de la protection appropriées et que leurs valeurs de preuve et d'information sont susceptibles d'être mises en évidence plus efficacement et plus facilement en ayant recours à des pratiques et des procédures normalisées.

<sup>5</sup> Les normes ISO 15930 – 1 à 15930 – 7 sont considérées et décrites comme des variantes de la norme ISO 15930.



Le format PDF/X a été mis au point aux Etats Unis par le Committee for Graphic Arts Technologies Standards (CGATS<sup>6</sup>) conformément aux exigences définies par la Digital Distribution of Advertising for Publications Association (DDAP).

Conçu pour permettre un échange facile mais spécifique pour les professionnels de l'édition et de l'impression, il fiabilise l'utilisation des fichiers graphiques, ce qui élimine les nombreux problèmes rencontrés dans les autres outils concernant les couleurs incorrectes, les polices manquantes et divers dysfonctionnements dans la création de fichiers graphiques (d'impression) pouvant engendrer une sortie incohérente sur les écrans et les imprimantes.

### **PDF/X-1**

Ce format date de 1999. Le PDF/X-1 : 1999 est une norme ANSI basée sur les spécifications de PDF 1.2. Seul l'espace CMJN<sup>7</sup> est accepté, il autorise un seul profil ICC<sup>8</sup> et les ressources (images, polices) doivent être incluses ou liées au document.

En 2001 il évolue légèrement (accepte le PDF 1.3) et le PDF/X-1 : 2001 devient, en tant que PDF/X-1a, la norme ISO 15930-1. Il est identique au PDF/X1, mais interdit les liens extérieurs et toute protection du document. Si la X-1 est très utilisée aux Etats-Unis, on lui préfère le plus souvent la X-3 en Europe. Le PDF/X1-a : 2003 est une mise à jour du PDF/X-1a : 2001.

### **PDF/X-2**

Le PDF/X-2, logiquement entre le PDF/X-1 et le PDF X-3, a été développé après le PDF/X-3. Il est toujours en cours d'élaboration mais il n'y est jamais fait référence dans les ouvrages techniques. Cette norme existe pourtant et figure au nombre des standards ISO (ISO 15930-2), mais n'a aucune présence sur le marché et n'est supportée par aucun logiciel professionnel.

### **PDF/X-3**

Il s'appuie sur le format de fichier d'Adobe<sup>®</sup> PDF 1.3, il est défini par la norme ISO 15930-3 et par la norme ISO 15930-6 depuis avril 2002.

---

<sup>6</sup> Le CGATS a été créé aux Etats-Unis en 1987 à la demande du comité Image Technology Standards Board (ITSB), une émanation de l'ANSI. Il a reçu son accréditation officielle en 1989.

<sup>7</sup> Le CMJN est un procédé d'imprimerie qui permet de reproduire un large spectre colorimétrique à partir des trois teintes de base (cyan, magenta, jaune) auxquelles on ajoute le noir.

<sup>8</sup> International Color Consortium (USA), qui a défini en 1993 les règles de gestion des couleurs pour harmoniser celles utilisées par les imprimantes avec celles utilisées pour l'affichage.





Le PDF/X-3 permet l'échange « aveugle » de documents, c'est-à-dire qu'il n'exige pas de connaître l'environnement de l'émetteur ou du récepteur. Il autorise la gestion de la couleur et accepte les espaces RVB<sup>9</sup>, CIELAB<sup>10</sup>, etc. C'est en fait la grande différence avec le PDF/X-1a qui lui n'accepte que le CMJN avec les tons directs. Cette version représente à ce jour le meilleur compromis entre liberté de création, fiabilité, possibilités de corrections simples et échange de données basées sur la gestion des couleurs.

#### **PDF/X-4**

Le format PDF/X-4 est décrit dans la norme ISO 15930-7 publiée en février 2008. Il s'appuie sur la version de structure de fichier PDF 1.6 de la société Adobe.

Cette version concerne l'échange complet de données d'impression et elle intègre aussi les possibilités d'échanges partiels de données en faisant référence à un profil externe dont la codification est représentée par la définition PDF/X-4p contenue dans la norme. Ainsi, tout échange de fichier ne nécessite plus la transmission obligatoire de l'ensemble des paramètres.

#### **3.3.3. PDF/E**

La lettre E pour : Engineering

Le PDF/E a été homologué par l'ISO en 2007. Il est utilisé par Acrobat<sup>®</sup> 9.0 pour les dessins techniques. Il est destiné à des utilisateurs comme les architectes, les professionnels du BTP, les centres de fabrication car il apporte une grande facilité d'utilisation et de restitution pour des dessins et des schémas complexes. Il est destiné à être utilisé comme un format d'échanges.

#### **3.3.4. PDF/UA**

Le sigle UA est donné à ce format de fichier pour : Universal Access. Il se veut être le format générique des PDF normalisés.

Le PDF/UA a été accepté comme référence de norme en mai 2008 par l'ISO. Il intègre toutes les possibilités de diffusion de documents contenant du texte, des dessins techniques, des révisions, des annotations et des schémas en 3D.

Ce format inclut tous les paramètres demandés et nécessaires pour s'interfacer avec des outils spécifiques à une utilisation des postes de travail par des personnes à handicap physique et à déficience visuelle.

---

<sup>9</sup> Format de codage des couleurs (rouge, vert, bleu).

<sup>10</sup> CIE = Commission Internationale de l'Eclairage. Elle a été fondée en 1913 à l'initiative de la France pour étudier les phénomènes de la vision dans son ensemble. CIELAB (ou CIE LAB) = Commission Internationale de l'Eclairage LABORatoire.



## 4. OUTILS LOGICIELS TRAITANT LE FORMAT PDF

Les outils sélectionnés dans ce paragraphe sont à ce jour les plus fiables au dire des techniciens. De nombreux autres disposent de fonctions de création, d’affichage, de transfert de fichiers et de documents au format PDF<sup>11</sup>.

### 4.1. OUTILS LOGICIELS DU MARCHE

Les principaux outils logiciels d’Adobe sont :

Adobe Reader<sup>®</sup> : affiche à l’écran tous les formats de fichier PDF (logiciel gratuit)

Adobe Acrobat<sup>®</sup> : transforme les fichiers bureautiques au format PDF, les sécurise, et permet leur partage

D’autres outils spécialisés sont décrits sur le site <http://www.adobe.com/fr/>

Les documents PDF créés à l’aide des outils du marché (autres que ceux d’Adobe) peuvent être ouverts dans Adobe Acrobat<sup>®</sup> et Adobe Reader<sup>®</sup> 4.0 et versions ultérieures.

Les outils de traitement de texte spécialisés (comme LaTeX) produisent des documents en PDF avec extraction du langage interne du fichier produit, donc sans possibilité de recréation du fichier original.

**Remarque :** les outils bureautiques (MS Office<sup>®</sup>, Open Office) ne créent pas directement de fichier PDF mais un fichier PS qui peut être repris par tout outil de création de fichier PDF (voir § 3.2).

### 4.2. OUTILS LOGICIELS DU MONDE LIBRE

#### 4.2.1. OUTILS LOGICIELS DE CREATION DE FICHIERS PDF

Parmi les nombreux logiciels existants on peut citer :

PDFCreator : transforme les fichiers PS en fichiers PDF

---

<sup>11</sup> A titre indicatif, il est possible de télécharger à partir du site de Microsoft<sup>®</sup> des outils qui permettent d’afficher, créer et indexer des fichiers au format XPS :  
<http://www.microsoft.com/downloads/details.aspx?displaylang=fr&FamilyID=b8dcffdd-e3a5-44cc-8021-7649fd37ffee>



HTMLDoc : génère automatiquement des fichiers PDF à partir de fichiers HTML (avec des images GIF, JPG, PNG)

Gymnast Outil : convertit les fichiers TXT en PDF

vPDFCreator : convertit les fichiers TXT en fichier PDF

#### **4.2.2. OUTILS LOGICIELS DE CREATION DE FICHIERS A PARTIR DU FORMAT PDF**

pdf2html : transforme des fichiers PDF en HTML

pdftotext : convertit les fichiers PDF en fichiers au format TXT. Certaines caractéristiques des caractères dans le fichier PDF ne sont pas conservées : caractères gras, soulignés, en italique.

Pstoedit : convertit les fichiers PS et PDF en formats graphiques DXF, HPGL, AI, WMH, EMF, Java, etc.

#### **4.3. OUTILS LOGICIELS DE VALIDATION DES FORMATS**

De nombreuses solutions de création de fichiers PDF s'avèrent être incomplètes du point de vue de la structure de fichier, ce qui peut générer une incompatibilité avec d'autres outils d'éditeurs.

De plus, pour toute structure de fichier faisant l'objet d'une norme, l'utilisateur doit disposer d'outils de contrôle de cette structure.

Pour ces motifs, outre Adobe® (depuis Acrobat® version 8), d'autres éditeurs proposent des solutions, gratuites ou payantes, pour vérifier la conformité de la complétude de fichiers au format PDF, c'est-à-dire qu'ils vérifient que la structure du fichier créé par une application informatique est en adéquation avec la norme à laquelle il fait référence<sup>12</sup>. Ces outils permettent donc de valider un logiciel (ou de réceptionner l'application) lorsqu'il y a incertitude sur les formats (cas des logiciels d'archivage, de création de PDF, etc.).

La société Sealsystems propose sur son site [www.sealsystems.fr/](http://www.sealsystems.fr/) un logiciel de contrôle de formats de fichier PDF avec l'outil PDF Longlife Suite. Il vérifie la conformité des fichiers PDF à la norme PDF/A.

PitStop Pro de Quartet Systems est un outil qui assure deux fonctions : contrôler l'adéquation du format d'un fichier PDF à la norme associée (essentiellement PDF/X) et sécuriser un document dans ce même format. Cet outil est téléchargeable sur le site : [www.quartet.fr/solutions\\_pdf/logiciels/pitstop\\_pro.html](http://www.quartet.fr/solutions_pdf/logiciels/pitstop_pro.html).

---

<sup>12</sup> Voir chapitre 3 sur les différents fichiers PDF et les normes associées.



## 5. L'APPORT DE LA VERSION 9 D'ADOBE® ACROBAT®

La nouvelle version 9 des outils mis sur le marché par la société Adobe apporte des évolutions dans la gestion d'informations graphiques. Les principaux points de cette évolution concernent l'intégration, dans le format des fichiers PDF 1.7, de contenus dynamiques comme la vidéo, l'animation graphique, l'audio.

Ces points d'amélioration sont :

Possibilité de **manipuler des fichiers 3D** convertis, c'est-à-dire que ces fichiers au format PDF peuvent être restitués sur les postes de travail sans disposer des outils de visualisation des applications natives.

**Conversion des fichiers cartographiques au format PDF** (fichiers provenant des outils SIG) et possibilité de réaliser automatiquement des calculs de distance entre deux points géographiques par l'intermédiaire d'une connexion automatique à un serveur de géolocalisation.

**Conservation de la sonorisation et de l'animation** des présentations sous Power Point lors de la conversion des fichiers au format PDF.

Intégration dans les outils Adobe® de l'**interactivité sur les fichiers PDF partageables** avec la possibilité, pour 3 personnes, d'ouvrir en simultané un document au format PDF, de le commenter et d'apporter des révisions.



## 6. LA SIGNATURE ELECTRONIQUE

Le premier objectif de la signature électronique d'un document est de certifier qu'il n'a pas été altéré depuis son origine.

La mise en place de la signature électronique doit être effectuée à moyen terme dans les systèmes de gestion documentaire lorsque l'archivage de documents sur support électronique est envisagé, comme le recommandent la législation et les normes internationales (Article 1316-4 du code civil).

Plusieurs solutions sont utilisées pour signer électroniquement les documents. La validité de ces solutions n'est reconnue légalement en France par les tribunaux que pour celles définies et décrites par les normes internationales.

### 6.1.LA SIGNATURE ELECTRONIQUE DEFINIE PAR LES NORMES INTERNATIONALES

Les aspects légaux de la signature électronique ainsi que les mécanismes qui la gèrent ont été définis au niveau international par les organismes de normalisation. Pour la France les aspects techniques en association avec les aspects légaux sont pilotés et décrits par la DCSSI.

La définition de la signature électronique donnée par la DCSSI est : « La signature électronique permet, à l'aide d'un procédé cryptographique, de garantir l'intégrité du document signé et l'identité du signataire. La signature électronique consiste en l'usage d'un procédé fiable d'identification garantissant son lien avec l'acte auquel elle s'attache ».

#### 6.1.1. ASPECTS LEGAUX DE LA SIGNATURE ELECTRONIQUE

La signature électronique est un mécanisme qui permet d'authentifier l'auteur d'un document électronique et de garantir son intégrité. L'analogie avec une signature manuscrite est présentée dans l'article 1316-4 du Code Civil.

Sur le plan légal, elle permet de répondre aux questions suivantes que se posent les juristes sur la validité des documents électroniques qui peuvent être produits en cas de besoins de fourniture de preuves tangibles :

- Est-ce ce que l'on accède au document original ?
- Le document est-il arrivé à bon port ? Est-il accepté par son destinataire ?
- L'émetteur est-il identifié ?
- Le document envoyé est-il l'original ou une copie ?



### 6.1.2. PROPRIETES DE LA SIGNATURE ELECTRONIQUE

Le mécanisme informatique de signature électronique doit présenter les propriétés suivantes :

- permettre au lecteur d'un document d'identifier la personne ou l'organisme qui a apposé sa signature,
- garantir que le document n'a pas été altéré entre l'instant où l'auteur l'a signé et le moment où le lecteur le consulte.

La signature électronique permet de garantir au document les états suivants :

- **Authentique** : L'identité du signataire doit pouvoir être retrouvée de manière certaine.
- **Infalsifiable** : La signature ne peut pas être falsifiée. L'émetteur ne peut se faire passer pour un autre.
- **Non réutilisable** : La signature n'est pas réutilisable. Elle fait partie du document signé et ne peut être déplacée sur un autre document.
- **Inaltérable** : Un document signé est inaltérable. Une fois qu'il est signé, on ne peut plus le modifier.
- **Irrévocable** : La personne qui a signé ne peut le nier.

Depuis mars 2000, la signature numérique d'un document a en France la même valeur légale qu'une signature sur papier (Loi N°2000-230).

## 6.2. PRINCIPE DES MECANISMES DE SIGNATURE

### 6.2.1. LES SIGNATURES NON RECONNUES LEGALEMENT

Le mode le plus couramment rencontré pour signer électroniquement un document est la copie par un outil bureautique (Word® ou Acrobat®) de la signature manuscrite de la personne, préalablement numérisée. Cette solution, qui présente de nombreux risques d'utilisation frauduleuse de la signature d'un tiers, n'a aucune valeur légale. Elle est équivalente à l'utilisation d'un tampon encreur pour signer un document.

D'autres procédés comme la numérisation de la page de garde signée manuellement puis fusion avec le reste du document et enregistrement au format PDF ne présentent aucune valeur légale, ni de présomption de preuve.

Les identifications d'étape de validation dans les systèmes de gestion de processus (type SAP par exemple) ne constituent pas non plus de preuve légale.



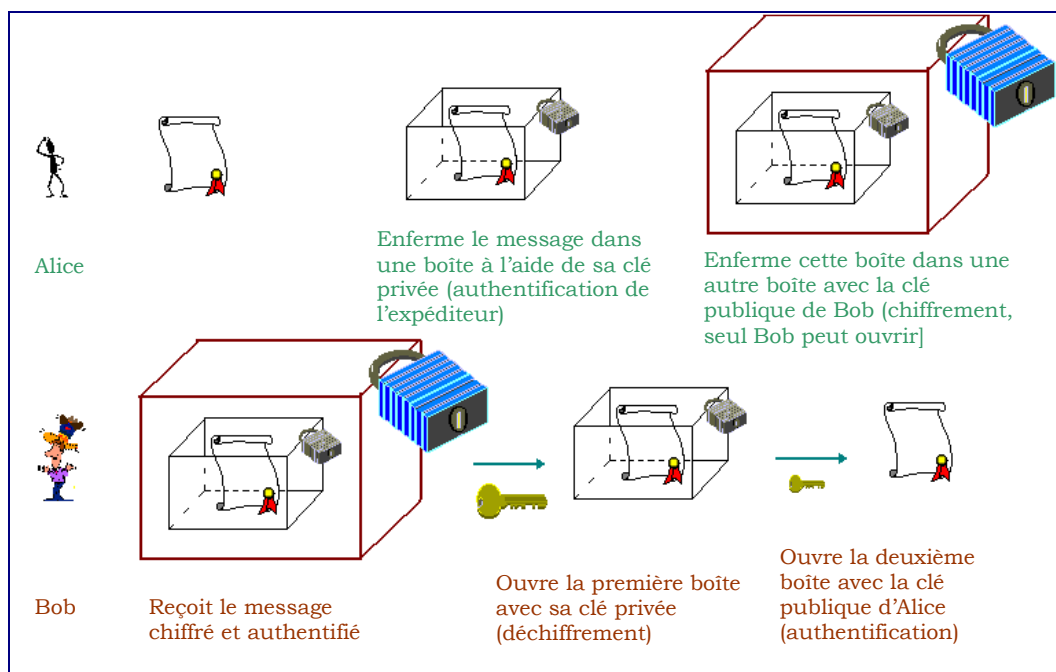
### 6.2.2. LES SIGNATURES LEGALES

Le principe de la signature électronique est l'application d'un algorithme au document électronique et à l'identifiant de l'émetteur (éventuellement du destinataire) pour obtenir une chaîne de caractères unique, non falsifiable, intégrée à la fin du fichier et vérifiable par un outil de contrôle. Cette opération est horodatée.

La garantie de non altération d'un document électronique est obtenue par l'apposition de la signature électronique de ce document et ensuite du passage d'un programme de contrôle qui recalculé la signature et la compare à l'initiale. Le niveau de fiabilité de l'algorithme permet de garantir que toute modification est identifiée par le programme de contrôle.

### 6.3. TRANSMISSION SECURISEE DE DOCUMENTS ELECTRONIQUES

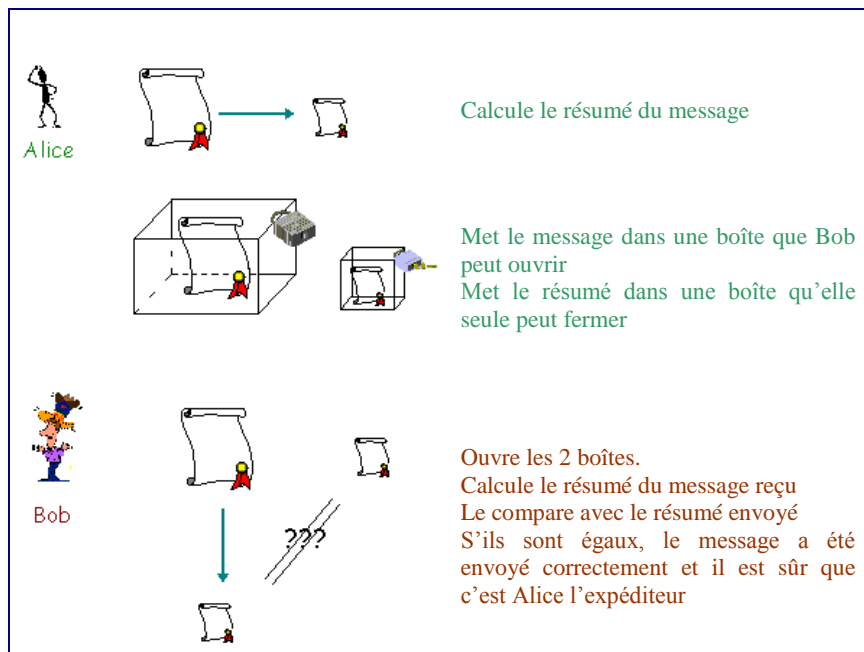
La signature électronique permet de sécuriser l'échange de documents entre individus ou applications informatiques (cas de la transmission par EDI) ; elle fait intervenir deux éléments secrets : la clé privée et la clé publique, pour rendre les signatures illisibles par un tiers non autorisé. Le schéma ci-dessous présente l'utilisation de ces deux éléments secrets dans un échange de documents entre Alice (émetteur du document) et Bob (récepteur).





On peut de plus rendre le document non lisible sauf par les personnes ou applications habilitées par une opération du cryptage qui consiste à appliquer un algorithme réversible sur le contenu du document pour le rendre incompréhensible. La clé de cryptage est détenue par l'émetteur et le destinataire. Le cryptage d'un document peut être intégré aussi bien dans la fonction de non altération d'un document archivé ou en complément de la signature électronique d'un document.

A ce transfert de documents signés, éventuellement cryptés, il faut ajouter la fonction de hachage qui calcule le résumé d'un texte. Ce résumé doit être à sens unique, pour éviter de pouvoir reconstituer le message initial à partir du seul résumé. Il doit être très sensible, c'est-à-dire qu'une petite modification du message entraîne une grande modification du résumé. En expédiant un message accompagné de son résumé (on dit aussi son « haché »), on peut s'assurer de l'intégrité du message, en recalculant le résumé à l'arrivée et le comparant à l'original.







## **7. RECOMMANDATIONS**

### **7.1.AVENIR DU FORMAT PDF**

Le format PDF est actuellement le format le plus utilisé dans le monde Internet et dans la conservation des documents. Il est toutefois incertain qu'à l'avenir il n'y ait pas deux autres formats qui lui soient préférés avec les mêmes finalités : le format HTML qui est en réalité un langage de présentation des documents au sens informatique et le format XPS qui est géré avec toute la puissance concurrentielle de Microsoft.

### **7.2.SECURISATION DU FORMAT PDF**

En natif, le format PDF a toutes les apparences d'un format sécurisé puisqu'il présente à l'écran l'image d'un document. En fait la sécurité d'un tel document n'est que relative puisque le format des fichiers est connu, donc modifiable à l'aide de programmes de mise à jour ou de modification.

Pour qu'un document au format PDF soit vraiment sécurisé, il doit être signé électroniquement comme le prévoient les normes et la législation.

La recommandation est d'utiliser le format PDF pour le stockage et l'archivage des documents produits, ce qui permet de garantir, le cas échéant, leur prise en compte par des outils de signature électronique des documents et de gestion d'archives électroniques, dès le stockage de leur version finale.

### **7.3.INTERPRETATION DES FICHIERS PDF**

La transformation d'un fichier PDF image en fichier interprétable se fait par appel à un outil de reconnaissance de caractères (ROC). Ce type de logiciel permet d'interpréter des images et de créer des caractères. La fiabilité actuelle de ces outils, pour des documents papier d'excellente qualité, est très proche de 99 % ce qui semble être satisfaisant ; cependant il faut relativiser ce pourcentage car une page au format Word® contient environ 2000 caractères et 1 % de deux mille est équivalent à 20 caractères erronés, donc à 20 mots potentiellement mal interprétés, ce qui n'est pas toujours acceptable.

La recommandation est de privilégier l'utilisation, dans les systèmes de GEIDE et d'archivage électronique, de formats PDF interprétables créés par des outils compatibles avec les suites logicielles utilisées et garantissant le respect des normes internationales.



## 7.4. UTILISATION DES DIFFERENTS FORMATS PDF

Les documents qui sont générés par les outils Adobe® sont dans le format propriétaire image ou interprété (PDF 1.7 actuellement), ceux créés par les outils de numérisation sont dans un format propriétaire image.

Les éditeurs de logiciels spécialisent leurs formats en fonction des destinations d'utilisation des documents :

PDF/A : pour l'archivage électronique des documents,

PDF/X : pour le monde de l'édition,

PDF/E : pour l'ingénierie documentaire,

PDF/UA : pour les échanges « universels » de documents.

Tous les formats PDF, qu'ils soient propriétaires ou normés, peuvent être transformés l'un dans l'autre avec des outils de transposition. Ces transformations de fichiers ne sont pas neutres : il y a systématiquement dégradation des informations de codage car tous les formats n'ont pas la même structure et ne contiennent pas les mêmes informations, ce qui diminue la qualité des fichiers en fin d'opération.

La recommandation est de limiter les transcodages successifs d'un fichier et, lorsque c'est possible, de reprendre le fichier initial pour le positionner dans la nouvelle structure.

## 7.5. TAILLE VARIABLE DES FICHIERS PDF CREEES

La taille des fichiers PDF en sortie d'outils bureautiques est très variable. Cette variation dépend principalement de deux facteurs :

- Premièrement, la police de caractères utilisée pour la création du fichier Postscript peut ne pas être reconnue dans Adobe® : ceci est par exemple le cas des fichiers produits par l'outil StarOffice<sup>13</sup> qui transmet au pilote des noms de polices non reconnus par Adobe®<sup>14</sup>. L'augmentation du volume par rapport au document original provient de l'intégration de la nouvelle police de caractères dans le fichier PDF.

---

<sup>13</sup> Suite bureautique concurrente de Microsoft qui a été développée par SUN.

<sup>14</sup> Certaines polices de caractères utilisées en bureautique sont la propriété d'éditeurs qui ont déposé la police. En conséquence les logiciels libres ne peuvent pas utiliser ces polices et doivent donc travailler avec des polices disponibles et non couvertes par des droits d'utilisation.



- Deuxièmement (et parfois simultanément avec le premier facteur) le volume dépend : de la façon dont l'application bureautique est interfacée au pilote d'imprimante et de celle dont est créé le fichier Postscript pour son impression.

La taille des fichiers PDF dépendra donc en grande partie des logiciels utilisés pour les créer.

La recommandation est de contrôler, avant l'installation des outils bureautiques, la taille et la structure des fichiers PDF générés.

### **7.6.LENTEUR DE L'IMPRESSION DE CERTAINS DOCUMENTS PDF**

Pour certains documents au format PDF, les temps d'impression de certaines pages sont très sont longs. La variation des temps d'impression des fichiers dépend :

- du poids relatif des pages, car il y a une complète dépendance entre ce poids et le temps de transit des informations dans le réseau entre le poste de travail et l'imprimante,
- de la codification des couleurs, du format des images, des polices de caractères qui peuvent demander des traitements dans l'imprimante et donc dégrader les temps d'impression.

Par ailleurs, l'imprimante doit convertir les fichiers PDF au format PS avant impression, ce qui peut être long, en particulier pour les formats image ; au contraire, une suite bureautique (type MS Office®) envoie directement des fichiers lisibles par l'imprimante reliée au poste de travail.

### **7.7.ARCHIVAGE ELECTRONIQUE ET FORMAT PDF**

Pour qu'un système d'archivage de documents électroniques soit en adéquation avec les règles qui définissent toutes les conditions de pérennité et de fiabilité, il devra initialement prendre en compte le facteur d'indépendance entre les informations archivées, les matériels (création, lecture) et les logiciels (création, contrôle, lecture et affichage, restitution, etc.).

La recommandation que nous donnons est de suivre, dès la création du document, la norme ISO 15 489.



## 8. REFERENCES BIBLIOGRAPHIQUES

Pour produire ce rapport, nous avons utilisé comme base les références documentaires suivantes :

- [1] Fiche produit : *Adobe® Acrobat® 9*. <http://www.adobe.com/fr/products/acrobat/>
- [2] Société Adobe, *Adobe donne l'intégralité de sa spécification PDF à l'ISO*, Communiqué de presse, 20 janvier 2007.  
[http://www.adobe.com/fr/aboutadobe/pressroom/pr/jan2007/adobe\\_pdf\\_iso.pdf](http://www.adobe.com/fr/aboutadobe/pressroom/pr/jan2007/adobe_pdf_iso.pdf)
- [3] Adobe Systems Inc., *PostScript Language Reference*, Third Edition, Addison-Wesley, Reading, MA, USA, 1999.
- [4] Norme ISO 19005: *Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)*, octobre 2005.
- [5] Adobe Systems Inc., *PDF Reference, sixth edition, Adobe® Portable Document Format Version 1.7*, Addison Wesley, Reading, MA, USA, 2006.
- [6] Catherine Dhérent (conservateur général), *Les archives électroniques : Manuel pratique*, Direction des Archives de France, février 2002.
- [7] Adobe Systems Inc., *Manuel de référence du langage PostScript (red book)*, Addison Wesley, Reading, MA, USA, 1992.
- [8] Adobe Systems Inc., *PostScript Language Tutorial and Cookbook (blue book)*, Addison Wesley, Reading, MA, USA, 1985.
- [9] Adobe Systems Inc., *PostScript Language Program Design (green book)*, Addison Wesley, Reading, MA, USA, 1988.
- [10] Pierre Blanc, *PostScript, l'essentiel*, Eyrolles, Paris, 1993.