

# Adobe® PDF, norme d'archivage

## Archivage de documents papier et électroniques au format Adobe PDF (Portable Document Format)

### Avantages et conditions de l'archives électronique

#### Défis de l'archivage moderne

Imaginez que, du jour au lendemain, vous ne pouvez plus pénétrer dans le bâtiment abritant vos dossiers les plus importants. Dans le monde actuel des affaires et de la politique, les entreprises comme les administrations commencent à s'intéresser de près à leurs processus ainsi qu'aux documents, archivés ou non, qu'ils produisent. Lorsque le principal support était le papier, archiver signifiait conserver du papier. Mais avec la création d'un nombre croissant de documents électroniques, quelle forme prend l'archivage ? Comment préserver de manière homogène une documentation mixte papier/électronique ? Comment éviter d'utiliser des documents papier ? Comment procéder pour qu'un document ait strictement le même aspect aujourd'hui et dans 30 ans ? Comment assurer une intégrité homogène des archives ?

L'apparition de l'ordinateur individuel dans l'entreprise a radicalement changé l'environnement de l'archivage. Jusqu'aux années 1990, la plupart des bureaux employaient encore un service de dactylographie ou de saisie sur traitement de texte, et conservaient leurs documents sur support papier dans des dossiers centralisés. Cependant, dès que l'informatique s'est généralisée la salle de classement centralisé a perdu toute utilité. Désormais, il appartient à chacun de créer, classer et gérer ses propres documents. Les entreprises et les administrations ont ainsi perdu le contrôle de ces documents.

#### Avantages des archives électroniques

La mise en place et le suivi permanent d'un système d'archivage électronique présentent de nombreux avantages. Par comparaison avec leur équivalent papier, les archives électroniques offrent un accès beaucoup plus facile à leur contenu, permettent de partager plus efficacement les informations et contribuent davantage à faire circuler le savoir. Désormais, les archives ne sont plus le domaine réservé de quelques personnes ayant une connaissance approfondie du système de classement. Pour permettre à un individu d'accéder aux informations contenues dans des archives électroniques, il suffit de lui octroyer les droits appropriés.

De plus, la forme électronique peut représenter des économies de coût non négligeables dans l'ensemble de la société. En effet, la gestion d'archives sur papier peut représenter un budget important que l'archivage électronique peut considérablement diminuer. Ainsi, le consultant Pricewaterhouse Coopers a publié une étude notoire selon laquelle une entreprise doit embaucher un nouvel archiviste à chaque douzième armoire-classeur qu'elle installe ! L'étude indique en outre que, dans le monde de l'entreprise, une personne passe 5 à 15 % de son temps seulement à lire des informations, mais jusqu'à 50 % à les rechercher.

#### TABLE DES MATIERES

1	Avantages et conditions de l'archivage électronique
1	Défis de l'archivage moderne
1	Avantages des archives électroniques
2	Conditions requises en matière de pertinence des documents
2	Conditions requises dans le contexte des fichiers PDF
4	Présentation générale du format Adobe PDF
4	Historique
4	Principes de base du format Adobe PDF
5	Adobe PDF en tant que format d'archivage
6	Normes relatives aux formats de fichier et aux métadonnées
6	Normes : de droit, de fait et autorisées
7	Rôle des métadonnées dans les archives électroniques
9	Processus d'archivage
9	Flux de travail de la création à l'archivage
10	Migration des archives aux fins de préservation
11	Avenir des archives numériques
11	Un projet international
11	Ressources

---

---

### Conditions requises en matière de pertinence des documents

Pour constituer des références valables permettant de retracer des actions et des décisions passées, les archives électroniques doivent être gérées de manière à préserver l'intégrité et l'authenticité de leur contenu. Pour cela, il est nécessaire d'adapter le programme de gestion des documents en place, et d'élargir la définition des documents de sorte qu'elle couvre non seulement le support papier, mais également la forme électronique. La création et la gestion d'un archivage électronique doivent être organisées au moyen de règles et de procédures, et d'une planification applicable à l'ensemble de la société. L'archivage doit en outre impérativement respecter les normes établies par l'entreprise.

Dans le domaine de la gestion des documents, la préservation du contenu, du contexte et de la structure des documents a toujours constitué une préoccupation majeure. Luciana Duranti, de l'UBC (University of British Columbia), mène des recherches dans le domaine des documents électroniques en utilisant comme fondement théorique la science diplomatique, dont elle applique les règles pour mesurer la fiabilité et l'authenticité du support électronique. Ses recherches démontrent que les organismes ont intérêt à appliquer les mêmes règles et procédures pour tous les documents, que ceux-ci se présentent sous forme de papier, de fichiers électroniques ou de microfiches. Cette homogénéité de traitement accroît en effet l'authenticité et l'intégrité des documents.

Les critères servant à déterminer la pertinence des documents conservés doivent répondre aux besoins commerciaux et juridiques propres à chaque organisme ainsi qu'aux réglementations et impératifs externes. A chaque entreprise ou administration correspondent donc des critères bien particuliers. Pour déterminer l'approche technologique la plus adéquate, il est indispensable de réaliser une analyse des risques approfondie avec le plein concours du service juridique. L'équipe d'évaluation doit être composée des membres suivants :

- Auditeurs et juristes : pour leur connaissance de la structure commerciale de l'entreprise, des procédures en vigueur et des principes et règles applicables aux documents ;
- Responsables de documentation et archivistes : pour leur connaissance des types de personnes autorisés à accéder aux documents, des motifs d'accès et de la durée d'accessibilité des documents ;
- Créateurs et utilisateurs des documents : pour leur connaissance de l'intérêt des documents et de leur valeur pratique.

### Conditions requises dans le contexte des fichiers PDF

Pour les responsables de documentation et les archivistes, un document doit être authentique, fiable, complet, intact et exploitable, et les systèmes électroniques permettant de le gérer doivent être aptes à protéger son intégrité dans le temps. Mais que signifient vraiment ces termes, en particulier lorsqu'ils s'appliquent aux fichiers PDF ?

- Authentique : il doit être possible de prouver qu'un document est réellement ce qu'il est censé être, qu'il a bien été créé ou envoyé par la personne prétendant l'avoir créé ou envoyé, et qu'il a bien été envoyé au moment indiqué. Ces garanties peuvent être fournies par l'emploi de métadonnées, c'est-à-dire de « données concernant les données ». Des métadonnées peuvent être incorporées aux fichiers PDF par un programme. La création, la réception et la transmission des documents doivent être contrôlées afin d'assurer que leurs auteurs sont dûment autorisés et identifiés. Il s'agit généralement d'une fonction du système général de gestion des documents électroniques, mais certains attributs des fichiers PDF, tels que les paramètres de sécurité, sont également conçus pour attester l'authenticité. Les signatures électroniques constituent un autre niveau d'authenticité supplémentaire applicable aux fichiers PDF.

---

---

Dans les transactions électroniques, un fichier PDF peut regrouper des données indiquant la personne ayant utilisé le système, à quel moment et pour quelles opérations, ainsi que le résultat de la transaction. Un programmeur astucieux peut utiliser ce type de fichier PDF pour extraire et conserver le maximum d'éléments de la transaction, et notamment la « présentation visuelle » complète de celle-ci. Cette présentation peut ensuite être restituée à l'utilisateur pour l'inviter à donner sa confirmation. Cela peut améliorer considérablement le score de l'entreprise aux tests d'admissibilité, car le document de la transaction présente celle-ci telle qu'au moment où l'utilisateur l'a confirmée comme étant acceptable.

- **Fiable** : il doit être possible d'avoir la certitude que le contenu d'un document est une représentation précise de la transaction dont il est la preuve. Le document doit être créé et enregistré en temps opportun par une personne directement concernée par l'événement, ou généré automatiquement par les traitements que l'entreprise applique habituellement pour mener une transaction. Cela s'applique particulièrement aux documents des transactions électroniques. Avec la spécification PDF, un intégrateur de systèmes peut automatiser l'enregistrement au format Adobe PDF de documents numériques directement depuis l'application source, que celle-ci soit une base de données, un programme de traitement de texte ou un tableur.
- **Complet et intact** : il doit être possible de protéger un document contre les modifications indésirables, mais également de surveiller et de suivre les annotations, ajouts et suppressions autorisés. Les règles et procédures de gestion de la documentation doivent préciser les ajouts et annotations pouvant éventuellement être apportés à un document après sa création, les circonstances dans lesquelles ces ajouts et annotations peuvent être autorisés, et les personnes habilitées à les réaliser. En principe, ces modifications affectent les métadonnées intégrées au document, et non le document même. Le système de gestion des documents dans lequel sont stockés les fichiers PDF confère généralement un niveau de sécurité élevé. Pour protéger les fichiers PDF, il est également possible d'utiliser des mots de passe et le cryptage. De plus, il existe des solutions de signature numérique et de PKI (public key infrastructure, infrastructure à clé publique), pour les documents PDF, fournies par des sociétés telles qu'Entrust, Inc. et VeriSign, Inc. Ces produits fonctionnent avec l'application Adobe Acrobat® en tant que plug-ins.
- **Exploitable** : il doit être possible de localiser, d'extraire, de restituer et d'interpréter un document, et de comprendre la suite d'opérations de sa création et de son exploitation, en remontant jusqu'à son origine si nécessaire. Grâce à la très récente fonction d'encapsulation de métadonnées XML dans les fichiers PDF, il est désormais possible de bénéficier de métadonnées de qualité « document » qui resteront lisibles et accessibles à long terme. De plus, les fichiers PDF peuvent être soumis à des recherches en texte intégral. Depuis des années, de nombreux fournisseurs de logiciels réputés, tels que Verity, Hummingbird et Convera, intègrent les fichiers PDF à leurs moteurs de recherche en texte intégral. Cela est dû au fait que la spécification et le Kit SDK des fichiers PDF sont publics, et qu'une bibliothèque PDF complète destinée au développement et à l'intégration de logiciels PDF est proposée pour un prix modéré.
- **Intégrité du système** : il doit être possible, pour assurer cette intégrité, de mettre en œuvre des mesures de contrôle telles que la surveillance des accès, la vérification des utilisateurs, la destruction autorisée, la sécurité et l'atténuation des effets des incidents majeurs. La fiabilité des systèmes est un facteur important d'intégrité, et les applications de gestion de documents excellent en ce domaine car elles sont conçues dans cet esprit. Toujours pour l'intégrité, il est important aussi d'utiliser comme format d'archivage standard un format de fichier bien documenté. La spécification PDF a été utilisée par une multitude de fournisseurs pour créer des applications uniques contribuant à assurer l'intégrité. L'intégration directe de solutions de signature numérique et de PKI fournies par d'autres éditeurs en est un bon exemple.

---

---

## Présentation générale du format Adobe PDF

### Historique

En 1985, Adobe participa à la création de ce qui fut alors baptisé « la révolution de la PAO » en lançant le langage de description de page Adobe PostScript.<sup>®</sup> Celui-ci permettait aux imprimantes de bureau d'effectuer le rendu de textes et d'images complexes. Ainsi, pour la première fois, la réalisation de publications haut de gamme n'était plus l'apanage des professionnels, mais se trouvait à la portée de n'importe quel utilisateur d'ordinateur. Il s'agit de l'une des applications décisives ayant encouragé les particuliers comme les entreprises à passer de la machine à écrire à l'ordinateur individuel.

En 1992, s'exprimant à propos des objectifs d'un projet de développement appelé Camelot, le cofondateur d'Adobe Systems Incorporated John Warnock déclarait : « Il n'existe aucun moyen universel pour communiquer et visualiser par voie électronique les informations imprimées... Ce qui manque maintenant aux entreprises, c'est un moyen universel qui permettrait de communiquer des documents entre des configurations de machine, des systèmes d'exploitation et des réseaux de communication très variés. »

La seule précision qui manque à sa déclaration est « dans l'avenir ». Le projet Camelot a mis au point la technologie appelée « PDF ». Celle-ci a permis au langage PostScript d'assurer le rendu de textes et d'images complexes à l'écran comme sur imprimante.

### Principes de base du format Adobe PDF

Bien qu'Adobe l'ait créé et continue à le faire évoluer en publiant de nouvelles versions, le format Adobe PDF est une spécification accessible au public. Il est courant de confondre PDF, qui est le format des données, avec Adobe Acrobat, qui désigne l'ensemble de logiciels commercialisé par Adobe pour créer, visualiser et améliorer les documents PDF. En 1993, la première spécification PDF était publiée, parallèlement au lancement des premiers produits Adobe Acrobat. Depuis cette date, Adobe met à la disposition du public, sur le Web, de nouvelles versions de cette spécification. A la date où nous publions le présent document, la dernière version de la spécification PDF est la version 1.4. Elle est disponible à l'adresse <http://partners.adobe.com/asn/developer/acrosdk/docs.html>. Toutes les révisions ayant fait l'objet de spécifications publiées sont rétrocompatibles : un ordinateur capable de lire la version 1.4 peut aussi lire la version 1.3, et ainsi de suite. Adobe ayant opté pour la publication de sa spécification PDF, le nombre d'outils de création, d'affichage et de manipulation conçus pour celle-ci par d'autres fournisseurs est de plus en plus important.

L'expression *Portable Document Format (Format de document transférable)*, ou *PDF*, a été choisie pour illustrer le fait qu'un fichier répondant à cette spécification peut être affiché et imprimé avec la même fidélité sur n'importe quelle plate-forme (UNIX,<sup>®</sup> Mac OS, Microsoft<sup>®</sup> Windows<sup>®</sup> et plusieurs appareils mobiles). Le document PDF reste identique sur toutes ces plates-formes. Il se compose d'une série de pages dont chacune contient du texte, des spécifications de polices, des marges, une mise en page, des éléments graphiques et des couleurs de fond et de texte. La présence de toutes ces informations garantit un rendu d'image précis, à l'écran comme à l'impression. Le fichier peut aussi comporter d'autres éléments tels que des métadonnées, des hyperliens et des champs de formulaire.

Pour que la spécification puisse être exploitée par des développeurs tiers, Adobe fournit à la fois le Kit SDK et la bibliothèque Adobe PDF. Ainsi, il est possible de développer des solutions complètes en dehors de la gamme de produits Acrobat, ou de modifier ces produits par la mise au point de plug-ins internes. Il arrive même que des développeurs n'utilisent la spécification que pour créer leurs propres visualiseurs ou créateurs PDF. Ces documents décrivent chaque aspect du format de fichier et la manière dont il peut être créé, lu et manipulé. En offrant un tel niveau d'assistance, Adobe a encouragé des sources variées à soutenir et à utiliser son format PDF.

### POUR EN SAVOIR PLUS

Les deux sites suivants constituent d'excellentes sources d'informations concernant Adobe Acrobat et les fournisseurs tiers : [www.pdfzone.com](http://www.pdfzone.com) et [www.pdfplanet.com](http://www.pdfplanet.com).

### Adobe PDF en tant que format d'archivage

Il existe pour l'archivage un choix important de formats et de technologies électroniques, notamment ASCII (pour le texte), TIFF, PDF et XML, ainsi que des formats tels que ceux des traitements de texte et des tableurs. Il est reproché à certains de ces formats d'être « propriétaires », et donc de ne pas offrir de garantie quant à leur existence à long terme. Un seul d'entre eux présente, par ses caractéristiques uniques, les conditions nécessaires pour assurer un affichage préservé sur une durée prolongée. PDF représente non seulement les données contenues dans le document, mais également l'aspect précis de ce dernier. Le fichier peut être affiché sans son application d'origine. En fait, dans dix ans et au-delà, les utilisateurs pourront le visualiser exactement tel qu'il était à sa création. De plus, l'ajout de métadonnées XML au fichier PDF garantit à la fois la fidélité et l'accessibilité de ce dernier. La spécification PDF étant à la disposition du public, les informations concernant le format de fichier resteront toujours dans le domaine public. Cela en fait donc un format de choix pour des archives électroniques. Les personnes atteintes d'un handicap peuvent en outre accéder aux informations au moyen de technologies d'assistance. Un mal-voyant, par exemple, peut utiliser un lecteur d'écran d'un fournisseur tel que Freedom Scientific, Dolphin Oceanic ou GW Micro pour verbaliser le texte. Pour cela, des balises sont incorporées dans la structure du fichier PDF. Ces balises peuvent être créées automatiquement depuis l'application d'origine, ou être entrées dans le cadre d'un traitement d'adaptation.

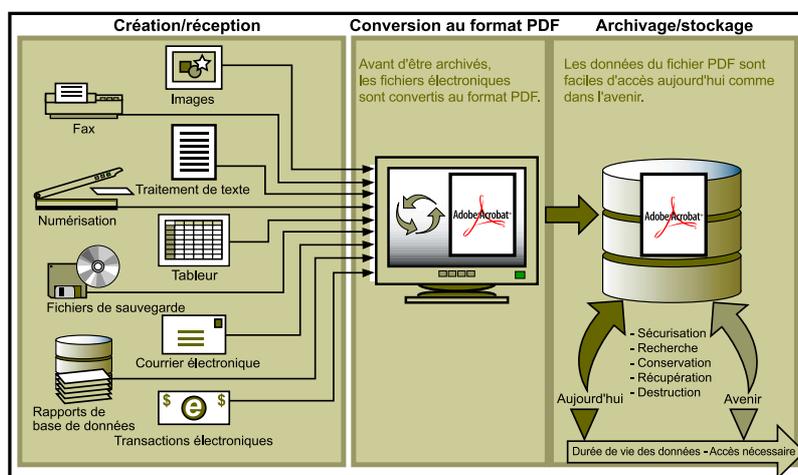
De nombreuses sociétés ayant recours à l'archivage électronique appliquent des procédures pour limiter la variété des formats des documents qu'elles reçoivent et conservent. Elles réduisent ainsi le nombre d'investigations portant sur les formats de fichier ainsi que de mécanismes sous-jacents nécessaires. Les archives nationales néerlandaises ont ainsi adopté des formats d'archivage électronique tels que PDF et XML. Le programme australien VERS (Victorian Electronic Record Strategy) utilise XML pour encapsuler les documents PDF avec des métadonnées normalisées. Le Bureau des archives publiques britanniques n'accepte que les formats PostScript, TIFF, SGML et PDF pour les fichiers devant être transférés aux archives.

Lorsqu'elles spécifient une technologie ou un format destinés à être employés par une grande variété d'utilisateurs, les organisations exigent de recevoir les fichiers de manière normalisée. Pour assurer la conformité et préparer leur communauté d'utilisateurs, elles doivent déterminer les caractéristiques que doit présenter un fichier PDF pour être estimé correct et satisfaisant pour ses besoins d'archivage. Elles peuvent exiger, par exemple, l'emploi limité de composants additionnels (incorporation de multimédia ou de JavaScript, par exemple), l'incorporation au document des polices utilisées, ou l'insertion par l'auteur ou par le destinataire du document de la totalité des métadonnées (voir en marge).

Actuellement, certains organismes d'Etat et représentants de l'industrie, ainsi que l'association AIIM International et le NPES/CGATS mènent un mouvement en faveur de la création d'une norme ISO autour du format Adobe PDF. Celle-ci serait particulièrement destinée à l'environnement de l'archivage. Ce nouveau projet, appelé PDF/A, aura de larges répercussions sur la conservation des documents à travers le monde. Pour plus d'informations sur PDF/A, consultez le site [www.aiim.org](http://www.aiim.org).

### Points à contrôler dans les archives PDF

1. Exigez l'incorporation des polices dans le document.
2. Normalisez l'archivage en imposant la présence de certaines métadonnées dans les fichiers, et veillez à ce que les utilisateurs n'en oublient aucune.
3. Veillez à ce qu'aucun fichier soumis ne soit crypté, ou associé à un mot de passe.
4. Déconseillez l'utilisation de code exécutable incorporé.
5. Standardisez la méthode de création de liens entre fichiers (par exemple, utilisez les liens relatifs si des fichiers font partie d'un ensemble).



---

---

## Normes relatives aux formats de fichier et aux métadonnées

### Normes de droit, de fait et autorisées

Les normes avalisées par un organisme de normalisation tel que l'ISO (Organisation internationale de normalisation) ou l'ANSI (American National Standards Institute) sont qualifiées de normes « de droit ». Les formats de données qui sont devenus des normes par la banalisation de leur emploi et de leur acceptation par les utilisateurs sont qualifiés de normes « de fait ».

Une norme de droit est développée sur une longue période, et doit être approuvée par chaque membre de l'organisme de normalisation ayant des intérêts dans son domaine d'application. Ces membres sont généralement des professionnels des secteurs concernés, des concepteurs de technologie, des ingénieurs et des experts en spécifications.

Les normes doivent être claires et concises, et ne laisser aucune place à l'interprétation. En effet, toute approximation peut entraîner des problèmes d'interopérabilité ou faire persister les problèmes qui existaient avant l'introduction de la norme. Voici quelques exemples de normes de droit :

- Z39.50 : Z39.50 désigne la norme internationale ISO 23950 : « Recherche d'information (Z39.50) : Définition du service de l'application et spécification du protocole » et la norme ANSI/NISO Z39.50.
- MARC 21 : MARC est l'acronyme de Machine-Readable Cataloging (Catalogage lisible par machine). Cette norme définit un format de données ayant pour origine une initiative de la Bibliothèque du Congrès américain datant d'il y a une trentaine d'années. MARC a pris le nom d'USMARC dans les années 1980, et de MARC 21 à la fin des années 1990. Il s'agit d'une norme décrivant le mécanisme selon lequel les ordinateurs échangent, utilisent et interprètent des informations bibliographiques, et ses éléments de données constituent le fondement de la plupart des catalogues de bibliothèques employés de nos jours.
- JPEG : JPEG est un mécanisme normalisé de compression d'image. C'est l'acronyme de Joint Photographic Experts Group, qui est le nom d'origine du comité ayant rédigé la norme.

Les normes de fait apparaissent pour satisfaire un besoin immédiat d'un secteur professionnel. Leur utilisation et leur popularité augmentent sous l'effet des exigences du marché. Elles sont généralement gérées par le groupe ou la société les ayant créées, et ne font l'objet d'aucun contrôle extérieur. Leur domaine d'application est généralement étroit, et elles servent un but spécifique. Elles s'imposent sur le marché et deviennent une norme parce qu'elles apportent la solution à des problèmes importants rencontrés par la profession. PostScript et PDF sont deux exemples de normes de fait.

Les normes autorisées recouvrent les normes réglementées et celles qui sont « recommandées pour assurer la compatibilité ». Les organismes qui autorisent des normes sont, par exemple, la FDA (Food and Drug Administration, Secrétariat américain aux produits alimentaires et pharmaceutiques) relativement aux dossiers d'autorisation de mise sur le marché de nouveaux médicaments. Les demandes peuvent être soumises par voie électronique, mais la FDA exige que la partie documentaire soit fournie sous la forme d'un fichier PDF doté d'attributs spécifiques, tels que des signets et des hyperliens. La FDA publie ses directives concernant les formats électroniques acceptables sur le site [www.fda.gov/cder/guidance](http://www.fda.gov/cder/guidance). Le projet australien VERS (Victorian Electronic Records Strategy) autorise le format PDF pour des documents qui sont ensuite encapsulés avec des métadonnées XML. Les organisations qui établissent leurs propres règles et procédures autorisent, en fait, des normes pour leur usage interne.

L'utilisation de PDF/X est un autre exemple de norme autorisée. PDF/X est une norme ISO concernant les secteurs de la presse et de la publicité. Dans ces secteurs, que les documents PDF soient créés ou reçus par un intervenant, il est important qu'ils soient toujours prêts à être produits immédiatement. Il s'agit d'un format PDF destiné aux professionnels des arts graphiques et qui

---

---

comprend un sous-ensemble spécifique de normes conformes à ISO 15930-1:2001. Pour plus d'informations sur PDF/X, consultez [www.pdfx.info](http://www.pdfx.info).

Le format PDF est également une norme autorisée du fait de l'utilisation, par le Bureau administratif des tribunaux et par la Cour des faillites du district sud de New York, d'une application Web permettant aux avocats de déposer par voie électronique les documents de déclaration de faillite. Le public peut visualiser ces documents à l'adresse suivante : [www.nysb.uscourts.gov](http://www.nysb.uscourts.gov). De nombreux cas de faillite notoires ont ainsi été traités au moyen du format PDF, qui a permis d'économiser un nombre incalculable d'heures de travail et de rames de papier. Ce sont ces documents PDF déposés par voie électronique qui constituent le dossier de chaque affaire. Ils ne font pas l'objet d'une confirmation sur papier. Selon Cecelia Morris, ancienne greffière de la Cour des faillites du district sud de New York, le principal avantage du dépôt d'un dossier juridique au format PDF est l'amélioration du service apporté aux parties concernées par l'affaire. Autrefois, l'accès aux fichiers appartenant au domaine public était long et difficile. Pour obtenir un document, la personne intéressée devait se déplacer jusqu'à la Cour, ou attendre que le document lui parvienne par courrier. Aujourd'hui, lorsqu'elle contacte la Cour, elle est invitée à consulter le site Web. Indépendant de la plate-forme et de l'application, le format PDF permet aux fichiers d'avoir l'aspect exact de leur original, quelle que soit la plate-forme sur laquelle ils sont consultés. « Pour les documents juridiques, la fiabilité des pages est un avantage décisif du format PDF », explique Cecelia Morris. « Avec ce format, vous avez l'assurance que chaque mot est identique au document officiel original. »

### Rôle des métadonnées dans les archives électroniques

La création d'archives électroniques nécessite celle de métadonnées. Le terme *métadonnées* nous vient du monde de l'informatique. Les professionnels de ce secteur utilisent ce concept de « données concernant les données » pour décrire les informations relatives à un objet. Les outils traditionnels de gestion des documents tels que les registres de dossiers, les couvertures de dossiers, les cartes de transfert, les thésaurus et les index fournissent des métadonnées. Ils aident les gestionnaires à contrôler et à gérer les documents. De plus, les outils de gestion de documents procurent d'importantes informations contextuelles concernant les personnes ayant utilisé les documents, l'usage qu'elles en ont fait, et à quel moment. Autrefois, les archivistes fournissaient des métadonnées supplémentaires en créant des index, des listes de dossiers et autres qui aidaient les personnes à localiser les documents et à comprendre leur organisation après leur transfert aux archives. Les métadonnées actuelles sont obtenues plus proactivement, les concepteurs de système cherchant à les extraire dès leur source, soit par un mécanisme automatique, soit par intervention de l'auteur des informations.

Ainsi, l'organisation Dublin Core Metadata Initiative utilise le système RDF (Resource Description Framework), car celui-ci permet aux plans de métadonnées d'être lus par les humains comme analysés par les machines, et de décrire plusieurs objets sans spécifier de détail supplémentaire. La seule exigence du « ciment » sous-jacent, XML, est que tous les espaces de noms doivent être définis. Une fois qu'ils sont définis, la personne insérant les métadonnées peut les employer autant que nécessaire. Les éléments de métadonnées de Dublin Core peuvent être inclus dans des fichiers PDF. L'exemple de métadonnées suivant, extrait d'un fichier PDF, identifie l'espace de noms Dublin Core (`xmlns='http://purl.org/dc/elements/1.1/'`) ainsi que trois éléments de métadonnées : créateur,

```
<rdf:Description about=''
  xmlns='http://purl.org/dc/elements/1.1/'
  xmlns:dc='http://purl.org/dc/elements/1.1/'>
  <dc:creator>Adobe Systems, Incorporated</dc:creator>
  <dc:title>Adobe Acrobat Help</dc:title>
  <dc:description>Adobe Acrobat</dc:description>
</rdf:Description>
```

---

---

titre et description. Les fichiers PDF contiennent, notamment, des métadonnées exprimées par les propriétés du document. Toute modification apportée dans la boîte de dialogue Acrobat Document Properties est répercutée dans les métadonnées. Celles-ci étant au format XML, elles peuvent être augmentées et modifiées au moyen de produits d'autres éditeurs. Un examen des métadonnées incluses dans les fichiers PDF met en évidence le fait que la spécification PDF intègre l'approche de l'initiative Dublin Core.

Les organismes doivent mettre en place une norme de métadonnées qui leur est propre pour spécifier le type d'information devant décrire l'identité, l'authenticité, le contenu, la structure et le contexte de leurs documents, et les exigences à respecter pour la gestion de ces derniers. Ces informations descriptives normalisées permettent alors de conserver à long terme des documents fiables, significatifs et accessibles qui satisferont les besoins commerciaux et pourront servir de preuves.

Divers projets internationaux existent en faveur de l'établissement de normes relatives aux métadonnées. Ils peuvent constituer de bons points de départ pour un organisme souhaitant normaliser ses pratiques dans ce domaine.

- Dublin Core Metadata Initiative, <http://dublincore.org/index.shtml>, août 2002
- Victorian Electronic Records Strategy Project, « VERS Metadata Scheme: Public Record Office Standard, PROS 99/007, Specification 2 », [www.prov.vic.gov.au/vers](http://www.prov.vic.gov.au/vers), août 2002
- Ministère américain de la Défense, « Department of Defense Directive 5015.2 Standard », [www.dtic.mil/whs/directives/corres/html/50152.htm](http://www.dtic.mil/whs/directives/corres/html/50152.htm), août 2002
- MARC 21, [www.loc.gov/marc](http://www.loc.gov/marc), août 2002

## Processus d'archivage

### Flux de travail de la création à l'archivage

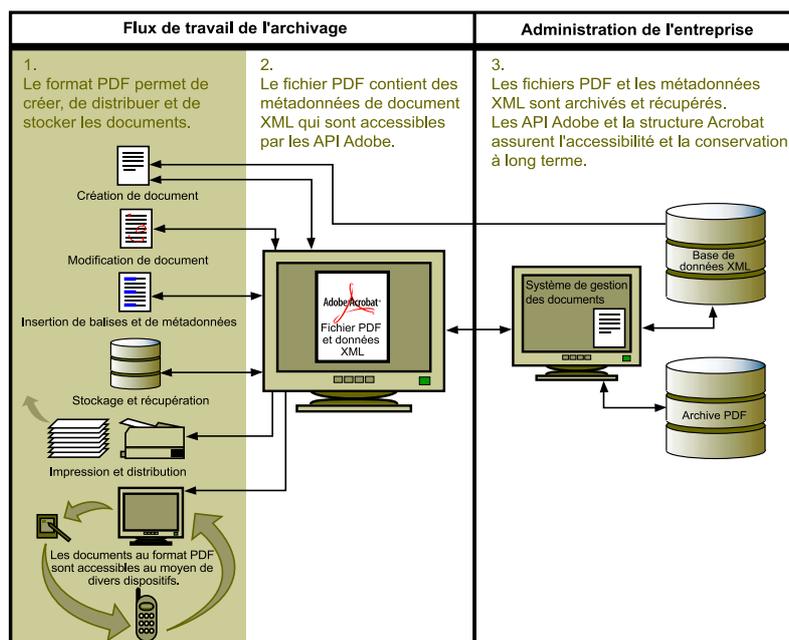
Dans une organisation où l'archivage est réalisé au moyen de dossiers sur papier, le flux de travail se résume à imprimer chaque document et à le classer. Cela suppose, naturellement, l'existence d'une salle de classement centralisée présentant encore suffisamment d'espace et munie d'un plan de classement compréhensible. Dans tout autre cas, l'archive papier n'est accessible qu'à la personne l'ayant classée. Cependant, passer d'un tel système à un archivage entièrement électronique ne se fait pas sans difficultés. Tout d'abord, il est nécessaire de mettre en place un ensemble complet de règles. L'organisme doit avoir une claire notion de ce qu'est un document d'archive, et de sa durée de conservation. Une bonne gestion de documents prévoit notamment la destruction en temps opportun.

Pour qu'un programme de gestion d'archivage soit efficace, il est nécessaire que des règles et des procédures soient énoncées, que les personnes soient formées à la mise en application de ces règles, et qu'un système de suivi de ces règles et procédures soit instauré et puisse être audité.

La plupart des règles et procédures de gestion d'archivage existantes ont été conçues pour les documents sur support papier. Pour se moderniser et commencer à archiver ses documents importants au moyen du format PDF, l'entreprise doit mettre à jour ses règles et procédures de sorte qu'elles soient adaptées à l'emploi accru de l'électronique tout en continuant à assurer la conservation d'informations juridiques et historiques pertinentes permettant de retracer les prises de décisions.

Pour mettre en place un archivage électronique, l'entreprise doit adopter un système de gestion de documents centralisé. Celui-ci doit pouvoir gérer à la fois les documents électroniques « d'origine », ceux convertis (par numérisation) du papier à l'électronique et ceux qui ne seront jamais électroniques (en raison de leur valeur, ou de frais de conversion trop élevés). La conversion au format Adobe PDF de documents qui ont toujours été électroniques peut être réalisée sur ordinateur de bureau, par traitement automatique ou par traitement serveur. Les logiciels Adobe Acrobat et Adobe Acrobat Distiller® Server peuvent, par exemple, exécuter cette opération. Pour convertir un document papier en fichier électronique, Adobe Acrobat Capture peut être employé pour numériser le document et le convertir au format PDF. Ce logiciel permet de numériser rapidement les archives papier d'une organisation.

Les fichiers PDF constituent un format d'archivage très pratique, car le texte qu'ils contiennent est accessible par le moteur d'indexation de recherche de texte intégral présent dans la plupart des systèmes de gestion de documents. Il est ainsi possible d'effectuer une recherche dans un document archivé en explorant à la fois ses métadonnées et son texte intégral. Il serait possible, par exemple, de retrouver des notes de service émises pendant une période comprise entre deux dates données et contenant les termes *décision* et *faillite*. Même les documents papier numérisés et convertis au format PDF peuvent faire l'objet de recherches. Pour cela, un moteur de reconnaissance optique des caractères doit être employé. Cette technologie identifie l'aspect du texte sur une page, et convertit l'image numérisée en texte reconnaissable.



---

---

### **Migration des archives aux fins de préservation**

Pour assurer la préservation d'informations numériques quelles qu'elles soient, il sera toujours nécessaire d'employer certaines techniques de conservation telles que les contrôles d'intégrité et les sauvegardes de sécurité. En matière de documents électroniques, des méthodes supplémentaires doivent être appliquées pour assurer que ces documents seront exploitables à long terme. Pour cela, deux approches sont possibles : l'émulation et la migration.

On désigne par « émulation » la reproduction de l'environnement technique requis pour manipuler des objets numériques anciens, par exemple l'exécution d'un programme DOS dans un système d'exploitation Microsoft Windows. La migration consiste à transférer systématiquement les données vers les configurations matérielles et logicielles les plus récentes. Chaque transfert doit être documenté et un contrôle doit assurer que ce transfert est complet. La migration est d'autant plus fiable que tous les documents électroniques se présentent selon un nombre limité de formats normalisés.

Le coût de la migration d'un document d'une version à l'autre d'un logiciel peut constituer l'un des principaux postes de dépense associés à la gestion du cycle de vie d'une archive électronique. L'opération peut consister à simplement ouvrir le document et à l'enregistrer dans le nouveau format. L'expérience démontre, cependant, que la migration n'est généralement pas si simple. L'ouverture de documents créés dans une version antérieure du logiciel peut entraîner des problèmes de mise en page, de numérotation des titres, de présentation d'objets graphiques, etc. Si ces problèmes sont parfois dus au logiciel, ils peuvent aussi résulter de la manière dont l'utilisateur a employé, ou tenté d'employer, une fonctionnalité du logiciel. Dans un tel cas, un utilisateur peut se voir dans l'obligation de consacrer du temps à reformater le document afin qu'il redevienne identique à l'original. S'il n'existe pas de version papier de l'original et que l'utilisateur ne dispose pas d'une copie de travail du document dans la version antérieure du logiciel, il peut même s'avérer impossible de restituer l'aspect initial. Le coût caché de ces opérations de migration réside dans la main-d'œuvre nécessaire pour veiller à ce que le document conserve son intégrité. C'est la raison pour laquelle tant d'organismes ont opté pour Adobe PDF comme format de conservation de leurs documents électroniques. Un fichier PDF représente la page imprimée telle quelle et, contrairement à un document enregistré dans un format de traitement de texte, il n'est pas susceptible de changer lorsqu'il est ouvert.

La spécification PDF étant accessible au public, votre organisme peut même créer un système de contrôle d'archive numérisant automatiquement les fichiers PDF entrants. Ces contrôles systématiques peuvent identifier les fichiers PDF incorrects, ou les fichiers non conformes à vos normes d'archivage. Vous évitez ainsi de procéder à des contrôles manuels coûteux sur tous les fichiers entrants.

# Avenir des archives numériques

## Un projet international

Dans le monde entier, la recherche se poursuit pour mettre au point des normes d'archivage numérique à l'épreuve du temps. Le format Adobe PDF occupe une place prépondérante à cet égard. La norme ISO PDF/A est un exemple d'initiative internationale indépendante qui reconnaît que l'intégrité de ce format en fait une norme de choix pour l'archivage électronique. Adobe s'est engagée à maintenir l'accessibilité publique de la norme PDF, et travaille en proche collaboration avec les organismes de normalisation pour que tout le potentiel de cette spécification soit exploité.

Cette facilité d'accès donne aux services d'archivage l'assurance que le format Adobe PDF existera encore pendant de longues années. N'importe qui peut, à n'importe quel moment et au moyen de n'importe quel matériel ou logiciel, créer des programmes d'accès à des archives électroniques. A vrai dire, il existe déjà une vaste communauté de développeurs qui créent des outils pour le format Adobe PDF. Cette communauté grandit chaque année.

Compte tenu des normes internationales, de l'abondance d'outils et de la technologie existant à l'heure actuelle, ainsi que de la nature publique d'Adobe PDF, celui-ci est le format tout indiqué pour les archives électroniques. Avec Adobe PDF, chacun a l'assurance de pouvoir accéder de façon rationnelle et permanente aux archives et aux documents au cours des années à venir.

## Ressources

« The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project » (« Conservation à long terme des documents électroniques : Résultats du projet InterPARES »)  
[www.interpares.org/book/index.cfm](http://www.interpares.org/book/index.cfm)

Digital Audio-Visual Preservation Prototyping Project (Projet de prototypage de conservation audiovisuelle numérique) [www.loc.gov/rr/mopic/avprot/avprhome.html](http://www.loc.gov/rr/mopic/avprot/avprhome.html)

U.K. Records Management for Central Government (Gestion des documents du gouvernement central du Royaume-Uni) [www.pro.gov.uk/recordsmanagement](http://www.pro.gov.uk/recordsmanagement)

Victorian Electronic Records Strategy Standards and Guides (Normes et guides du VERS)  
[www.prov.vic.gov.au/vers/standards/standards.htm](http://www.prov.vic.gov.au/vers/standards/standards.htm)

Adobe Systems France • Tour Maine Montparnasse, 33 Avenue du Maine, BP 14, 75755 Paris Cedex 15, France • [www.adobe.fr](http://www.adobe.fr), [www.adobe.com](http://www.adobe.com)

Adobe, le logo Adobe, Acrobat, Acrobat Capture, le logo Adobe PDF, Distiller, PostScript et "Tools for the New Work" sont des marques ou des marques déposées d'Adobe Systems Incorporated aux Etats-Unis et/ou dans d'autres pays. Mac est une marque d'Apple Computer, Inc., déposée aux Etats-Unis et dans d'autres pays. Microsoft et Windows sont des marques ou des marques déposées de Microsoft Corporation aux Etats-Unis et/ou dans d'autres pays. UNIX est une marque déposée de The Open Group. Toutes les autres marques citées sont la propriété de leurs détenteurs respectifs. © 2003 Adobe Systems Incorporated. Tous droits réservés. Printed in France. 95001214 1/03



Tools for the New Work™