

Guide Méthodologique

Le format de fichiers PDF

Table des Matières

CONTROLE DU DOCUMENT	3
AUTEURS & VERSIONS	3
LISTE DE DISTRIBUTION POUR CETTE VERSION	3
SURVOL DU FORMAT PDF	4
INTRODUCTION	4
LES ORIGINES DU PDF : POSTSCRIPT	5
ORIGINES	5
DESCRIPTION	5
POURQUOI LE PDF ?	8
HISTORIQUE	8
CARACTERISTIQUES	8
FONCTIONNALITES	14
STRUCTURE DU FICHIER	14
MODELISATION DE L'IMAGE	14
TABLEAU RECAPITULATIF DES FONCTIONNALITES	24
INTRODUCTION AUX NORMES BASEES SUR LE PDF	25
SCHEMA DES LIENS ENTRE LES VERSIONS DE PDF	26
ADOBE ACROBAT	27
CONCLUSIONS	27
GLOSSAIRE GENERAL	29
GLOSSAIRE PDF	30
BIBLIOGRAPHIE	34
LES VERSIONS DE PDF	35
PDF v1.0	35
PDF v1.1	38
PDF v1.2	41
PDF v1.3	44
PDF v1.4	48
PDF v1.5	52
PDF v1.6	56
PDF v1.7 (ISO 32000-1)	59
PDF v1.7 NIVEAU D'EXTENSION 1	61
PDF v1.7 NIVEAU D'EXTENSION 3	62
PDF v1.7 NIVEAU D'EXTENSION 5	65
PDF v2.0 (ISO 32000-2)	66
LES NORMES PDF	70
PDF/A	71
PDF/E	83
PDF/UA	91
PDF/VT	95
PDF/X	98
REFERENCES	109

Contrôle du Document

Auteurs & Versions

Nom de l'Auteur	Organisme	Version	Date publication	Changements par rapport à la Version précédente
Nick Parker	Numen	0.1		Version initiale
Michel Bouloc	CINES	0.2	22/05/2012	Modifications
Olivier Rouchon	CINES	1.0	08/08/2012	Changement de feuille de style, mise en page, corrections mineures – version diffusée

Liste de distribution pour cette version

Nom	Titre	Rôle
Olivier Rouchon	[CINES] Responsable Département Archivage et Diffusion	Rédacteur
Michel Bouloc	[CINES]	Relecteur
Michel Jacobson	[SIAF]	Relecteur
Françoise Banat-Berger	[SIAF]	Relecteur

Survol du format PDF

Introduction

Les fichiers PDF sont omniprésents. Les utilisateurs réguliers de l'informatique ont sûrement ouvert et consulté un fichier PDF. Ce format existe depuis 1993 et il est devenu norme ISO en 2008.

Durant ses 18 ans d'existence, le format PDF a évolué avec sept révisions. Plusieurs normes ISO ont été élaborées pour préciser comment utiliser les fichiers PDF dans des domaines spécifiques. Comment se retrouver dans toutes ces versions, savoir quelle norme ou quelle version utiliser, choisir les fonctionnalités significatives dans un fichier PDF : maîtriser toutes ces informations devient difficile aujourd'hui.

La première partie de ce rapport a pour but d'identifier les différentes fonctionnalités pouvant être intégrées dans un fichier PDF. Il y en aura une description succincte de chacune entre elles. Le lecteur pourra ainsi avoir un aperçu des différentes possibilités du format PDF.

La deuxième partie décrira chaque version et la troisième partie chaque norme. Les différences seront expliquées pour aider le lecteur à faire les bons choix des versions et des normes à utiliser.

Ce document n'a pas pour but de fournir les détails techniques de mise en œuvre du format PDF. Les spécifications sont librement disponibles. Son objectif principal est d'étudier l'ensemble de ces versions et normes pour aider à comprendre les différentes spécificités afin d'effectuer le meilleur choix.

Les origines du PDF : PostScript

Une étude, même superficielle, du format PDF ne peut pas être complète si on ne commence pas avec une présentation du langage PostScript. Ce langage était le précurseur de PDF.

PostScript est un vrai langage de programmation (dit Turing-complet), optimisé pour faire la description de pages.

Origines

PostScript trouve ses racines dans un langage appelé « Design System » développé, à partir de 1976, par l'entreprise Evans & Sutherland Computer Corporation. C'était le résultat d'un projet de recherche qui avait pour but de représenter des bases de données de graphiques à trois dimensions, avec un langage interprété.

La majorité des idées pour « Design System » venait de John Gaffney. A cette époque, il travaillait pour John Warnock. Lorsque John Warnock a pris un poste à Xerox Palo Alto Research Center en 1978, il a réécrit le langage (maintenant appelé « JaM »). Ce langage a été utilisé à la fois pour la conception de circuits intégrés par Martin Newell, pour des expérimentations sur l'impression et pour les arts graphiques par John Warnock.

En 1982, John Warnock a créé l'entreprise « Adobe Systems » avec Chuck Geschke. Il a proposé une troisième version du langage, maintenant appelé « PostScript », essentiellement utilisé pour la description de pages imprimées en deux dimensions.

La première machine à incorporer un interpréteur PostScript était l'Apple LaserWriter, sortie en 1985. Ces imprimantes étaient, avec PageMaker sur le Macintosh, à l'origine de la révolution PAO des années 1980.

Description

PostScript est un langage interprété, basé sur une pile et utilisant la notation polonaise inverse.

PostScript comprend des opérateurs décrivant la séquence des opérations, mais aussi des opérateurs spécifiques à la description des pages.

Certains de ces opérateurs spécifiques sont :

- **Définitions de chemin** : des formes arbitraires peuvent être construites à partir des lignes droites, des arcs et des courbes. Ces formes peuvent contenir des sections discontinues et des trous ;
- **Opérations de dessin** : une forme peut être utilisée pour tracer son contour avec une ligne de n'importe quelle épaisseur, remplie avec n'importe quelle couleur, ou utilisée pour détourner d'autres formes ou des images rastérisées ;
- **Le texte** : le texte est intégré complètement avec des formes graphiques. Les caractères de texte peuvent être utilisés comme des formes par tous les opérateurs graphiques ;
- **Images rastérisées** : des images rastérisées peuvent être imprimées à n'importe quelle résolution et manipulées pour donner une variété d'effets graphiques ;
- **Coordonnées et transformations** : PostScript comporte des fonctionnalités pour appliquer des transformations linéaires aux systèmes de coordonnées, pour réaliser des déplacements, des mises à l'échelle, des rotations, des réflexions et des déformations. Ces transformations peuvent s'appliquer à des formes, à du texte et à des images rastérisées.

Exemple

Un fichier PostScript est simplement un fichier texte. L'exemple ci-dessous montre un programme PostScript qui imprime le texte « Texte sur la page » en police Helvetica sur la page :

```
%!PS
/Helvetica findfont 10 scalefont setfont
100 100 moveto
(Texte sur la page) show
showpage
```

La première ligne est simplement un commentaire pour identifier le fichier comme étant du PostScript.

La deuxième cherche la police Helvetica et l'applique avec un corps de 10 points.

La troisième ligne positionne le point courant à une distance de 100 points horizontalement et verticalement du coin inférieur gauche de la page.

La ligne suivante affiche le texte.

La dernière ligne imprime la page.

Évolution du langage

Le langage PostScript proprement dit a évolué avec trois versions majeures. Il y a eu des versions mineures introduites régulièrement afin de suivre des évolutions de technologie.

PostScript Niveau 1

La version de base du langage introduit en 1984. Cette version ne comportait pas la gestion de couleur CMJN nécessaire à l'impression professionnelle, uniquement les couleurs RVB et TSV (teinte, saturation, valeur).

PostScript Niveau 2

Introduit en 1991, cette version a apporté trois évolutions principales :

- **couleur** : modélisation CMJN et support pour les images rastérisées en couleur ;
- **polices composites** : pour gérer les polices de grande taille ou d'encodage spécifique ;
- **Display PostScript** : des extensions pour permettre l'utilisation de PostScript sur les écrans d'ordinateur.

PostScript 3

Introduit à la fin de 1997, les principales nouveautés de cette version sont :

- des filtres de compression et de décompression ;
- l'ajout d'espaces de couleur pour l'utilisation de couleurs indépendantes du périphérique et des couleurs composites ;
- amélioration du nombre de niveaux pour pouvoir réaliser des dégradés plus lisses ;
- un mécanisme pour l'encodage des caractères des polices sur plusieurs octets (CID-keyed) ;
- support pour les formats de police CFF, Chameleon, TrueType et des polices bitmap ;
- support pour le grossi/maigri automatique dans le RIP.

La version 3 était la première version à vraiment être utile pour l'impression professionnelle en couleur.

Adobe continue à améliorer son interpréteur PostScript et les produits PostScript 3 d'Adobe permettent actuellement l'impression directe de fichiers PDF jusqu'à sa version 1.5. Grâce à ses produits, il est ainsi possible d'imprimer des images qui utilisent les fonctionnalités de transparence de PDF, alors même que le langage PostScript ne permet pas de les décrire.

Inconvénients

Bien que PostScript ait eu une influence considérable et qu'il était possible de décrire pratiquement n'importe quelle page avec, il restait malgré tout des inconvénients qui ont fait qu'il a été supplanté par PDF.

Le premier problème était la difficulté de visualiser une page PostScript sur un écran d'un ordinateur. Écrire un système de visualisation revenait à écrire un interpréteur complet du langage. Adobe n'a jamais créé un interpréteur pour écran, sauf sur les systèmes qui utilisaient *Display PostScript* (une variante pour écran) comme langage de base de leur affichage (comme les postes de travail Sun par exemple). Ces machines étaient toujours chères, et *Display PostScript* n'a jamais atteint une masse critique. Les ordinateurs de bureau n'étaient pas suffisamment puissants pour ce système.

La solution *GhostScript* (un interpréteur PostScript libre) permettait d'afficher des pages PostScript à l'écran, mais c'était un processus lourd et trop lent sur les machines de l'époque.

Le deuxième problème était que les pages d'un fichier PostScript pouvaient être interdépendantes ; c'est à dire que l'affichage de la deuxième page pouvait dépendre d'éléments définis à l'intérieur de la première page. Il était donc impossible de manipuler les pages de façon indépendante, et pour afficher une page il fallait avoir interprété toutes les pages précédentes.

Enfin, un système de commentaires a été introduit, qui permettait d'ajouter une structure au fichier PostScript au travers d'un ensemble de règles préconisées, mais ces règles n'étaient pas obligatoires, et rien dans le langage ne permettait de les contrôler.

C'est essentiellement à cause de ces trois problèmes que le PDF a été développé.

Pourquoi le PDF ?

Historique

Avec le constat de la lourdeur du langage PostScript et le fait, qu'à l'époque, les ordinateurs de bureau n'étaient pas suffisamment puissants pour faire fonctionner un interpréteur PostScript convenablement, John Warnock a décrit en 1991 une façon de normaliser les fichiers PostScript ces fichiers (appelés IPS ou « Interchange PostScript ») permettent de créer facilement des programmes capables de les visualiser. Au début cette normalisation se faisait par un interpréteur PostScript. John Warnock envisageait que PostScript serait toujours la solution à long terme, et que IPS ne serait qu'une solution temporaire en attendant que les ordinateurs deviennent plus puissants.

Finalement, le format IPS a évolué vers le format PDF.

Bien que propriétaire de la société Adobe, le format des fichiers PDF a toujours été publié. La première version du *PDF Reference Manual* a été éditée par la maison d'édition Addison Wesley, en 1993.

Le format des fichiers PDF a évolué régulièrement depuis cette première version, mais toujours avec l'objectif que les anciens fichiers PDF puissent être lus avec les nouveaux visualisateurs, ainsi que les nouveaux fichiers sur les visualisateurs de version antérieure.

La deuxième partie de ce rapport va examiner en détail chacune des évolutions. Ceci afin d'indiquer à quel moment chaque fonctionnalité a été introduite. Cela permettra de bien comprendre les différences entre les versions et d'identifier la version minimale adéquate pour utiliser une fonctionnalité spécifique.

Bien que les fonctionnalités permises dans un fichier PDF aient évoluées avec le temps, le modèle de base de la description de l'image est resté quasiment inchangé depuis le langage PostScript.

Adobe garde sur son site un certain nombre d'archives sur l'historique du format PDF :

http://www.adobe.com/devnet/pdf/pdf_reference_archive.html

Caractéristiques

Dès la première version de PDF, les caractéristiques énoncées par Adobe pour le PDF étaient les suivantes :

Modélisation de l'image utilisant le modèle PostScript

Une description de page PDF utilise la même façon de décrire une page qu'un programme PostScript.

Il dessine une page en plaçant de l'encre sur des régions sélectionnées :

- les formes dessinées peuvent être des caractères, des régions définies par des combinaisons de traits et de courbes, ou des images rastérisées ;
- l'encre peut être de n'importe quelle couleur ;
- n'importe quelle forme peut être rognée par une autre forme afin que les parties de la forme originale à l'intérieur de la deuxième apparaissent ;
- chaque marque ajoutée sur la page obscurcit complètement les marques déjà présentes.

Ce principe de description de la modélisation de l'image est resté exact avec les versions subséquentes de PDF. Seule l'introduction de la transparence, avec la version PDF 1.4, a ajouté des possibilités dans PDF qui n'existe pas dans PostScript.

Portabilité

Le format originel de PDF ne comportait que des caractères 7-bit (ASCII) pour permettre une transmission fiable sur les systèmes de l'époque (où le huitième bit était souvent perdu). Cette restriction a été enlevée par la suite. PDF donne une définition claire de l'ordre des octets et de l'utilisation des fins de ligne, ainsi le format de fichier ne dépend pas du système d'exploitation.

Le format PDF peut ainsi être créé et lu sur tout type de système sans souci de compatibilité.

Compression

Dès l'origine, le format PDF a permis une compression de son contenu. Ce qui permet de réduire sensiblement la taille des fichiers par rapport à des fichiers non-compressés.

Indépendance des polices

L'intention originelle de PDF était de ne pas inclure les polices dans les documents, mais d'inclure un descripteur moins gourmand en taille d'octet. Ceci permettait au système qui lisait le fichier, de substituer la meilleure police possible lorsque la police d'origine n'était pas disponible. Il était accepté que les polices casseaux nécessitent un traitement spécifique. Ces polices étaient incluses en version compressée pour que les symboles soient restitués correctement.

L'objectif originel de créer des fichiers aussi faible en taille que possible a été remplacé, plus tard, par l'objectif de garder un rendu exact. Ceci en encourageant l'inclusion de toutes les polices dans un fichier PDF, mais en conservant aussi la possibilité de substitution.

Génération de fichier en un passage

Il existe de nombreux formats de fichiers indexés qui nécessitent plusieurs passages lors de l'écriture d'un document. En effet il y a nécessité de créer des liens entre les parties différentes du document. Ces liens ne peuvent être créés qu'une fois la taille de chaque partie connue.

Le format de PDF a été conçu pour éliminer ce processus. Bien qu'il existe des liens dans le fichier, il est toujours possible d'écrire les différentes parties pour que ces liens indiquent ce qui a déjà été écrit et dont la position est connue.

L'index général des positions des objets dans le fichier est positionné à la fin de celui-ci.

L'écriture d'un fichier PDF est donc rapide et peut être faite sur un système avec des ressources limitées.

Accès direct à chaque page

Un des problèmes de PostScript était que chaque page pouvait dépendre des pages précédentes. PDF enlève cette possibilité. Chaque page est indépendante et un index de pages permet un accès direct à chaque page sans besoin de traiter les autres.

Mise à jour incrémentale

Le format de PDF permet qu'on puisse modifier un fichier PDF de façon à seulement ajouter à la fin du fichier existant. Cette possibilité permet l'enregistrement rapide des modifications.

La conséquence de cette fonctionnalité est que, si on supprime 50 pages d'un fichier PDF de 100 pages, la taille du fichier augmente au lieu de se réduire. En effet, l'information de la suppression des pages sera enregistrer à la fin du fichier, au lieu d'une réécriture complètement du fichier.

L'avantage de ce système est l'enregistrement rapide des modifications et la possibilité de tracer l'historique du fichier. Le désavantage est qu'il est mal compris par les utilisateurs et il en résulte des fichiers beaucoup plus grands que nécessaire.

Cette fonctionnalité existe dans PDF depuis l'origine. Dans Acrobat, par exemple, lorsque l'on utilise l'option « Fichier/Enregistrer », il y a ajout des modifications à la fin du fichier. Et lorsque l'on utilise « Fichier/Enregistrer sous » (même sans changer le nom du fichier), le fichier est réécrit complètement. Avec Acrobat, il est donc nécessaire d'utiliser la deuxième option si l'on veut des fichiers de taille optimale.

Extensibilité

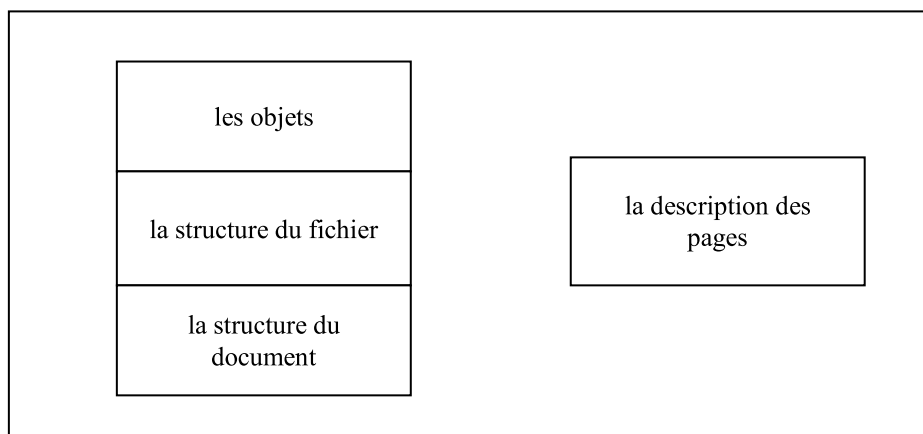
Le format d'un fichier PDF a été pensé dès l'origine pour permettre des extensions. N'importe quel logiciel est donc toujours capable d'ajouter ses propres extensions dans un fichier PDF qui seront ignorées par les autres lecteurs PDF.

Grâce à cette structure, cela a permis l'ajout de fonctionnalités, au fur et à mesure des versions, sans perdre la compatibilité de lecture avec les fichiers et les lecteurs originaux.

Structure des fichiers

Un fichier PDF est essentiellement un fichier texte qui est lisible avec un éditeur de texte. Le contenu de base du fichier PDF sera en binaire, uniquement avec l'application de la compression ou du cryptage de tout ou une partie du fichier. Les éléments inclus (polices, images, fichiers embarqués, ...) seront aussi en binaire.

Un fichier PDF peut être décrit avec quatre niveaux de détails :



Les objets

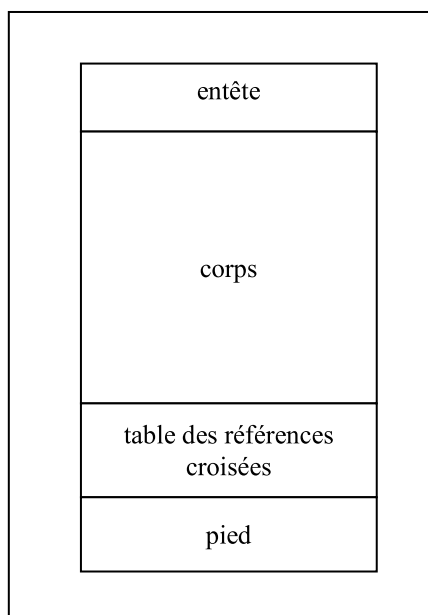
La plupart des éléments d'un fichier PDF sont décrits avec un nombre limité d'objets de type différents :

- la valeur **booléenne** qui indique une valeur vraie ou fausse. Exemples : `true`, `false` ;
- les **numéros** : valeurs entières ou réelles. Exemples : `123`, `-0.23` ;
- les **chaînes de caractères** : on utilise des parenthèses pour délimiter la chaîne. Exemple : `(chaîne de caractères)` ;
- les **noms** utilisés pour identifier d'autres objets PDF. Exemple : `/Font34` ;
- les **matrices** unidimensionnelles (arrays) qui peuvent contenir un nombre arbitraire de différents types d'objets dans un ordre précis. Exemple : `[ 549 false (texte) ]` ;
- les **dictionnaires** qui permettent de faire correspondre des noms avec des objets de type arbitraires. Les dictionnaires sont les briques de base d'un fichier PDF. Exemple : `<< /Type /Exemple /Version 1.0 >>` ;
- les **flux** (streams) qui peuvent contenir n'importe quelle séquence d'octets.

En plus des objets directs (où un objet apparaît à l'endroit où il est utilisé dans le fichier), il existe des objets indirects qui sont stockés de façon à pouvoir les trouver par un accès direct. Ce mécanisme permet de réutiliser un objet plusieurs fois, mais sert aussi comme structure de base du fichier PDF puisque tous les éléments importants du fichier sont stockés comme des objets indirects. Il est en général possible de remplacer un objet direct par un appel à un objet indirect si nécessaire.

La structure du fichier

Un fichier PDF comporte quatre parties :



- l'**entête** — une simple ligne de texte qui indique que le fichier est un fichier PDF et fournit sa version : %PDF-1.5 ;
- le **corps** — une séquence d'objets indirects utilisés pour décrire le contenu du fichier ;
- la **table des références croisées** — permet un accès direct aux objets indirects contenus dans le corps du fichier ;
- le **pied** du fichier — permet de trouver très vite la table des références croisées et un certain nombre d'autres objets qui sont importants pour la lecture du fichier.

Comme déjà mentionné, le format PDF permet des modifications incrémentales en autorisant l'ajout des modifications à la fin du fichier sans toucher le contenu existant. Pour faire cela, après le pied original il est possible d'ajouter :

- un corps qui contient des nouveaux objets qui s'ajoutent à, ou remplacent, des objets existants ;
- une table des références croisées partielle ;
- un nouveau pied de fichier.

Guide Méthodologique

Le format PDF

entête originale
corps original
table originale des références croisées
pied original
corps ajouté
table des références croisées ajoutée
pied ajouté

Il n'y a pas de limite sur le nombre de mises à jour qu'il est possible d'ajouter de cette façon, mais la lecture du fichier va devenir plus lente, et la taille du fichier plus grande à chaque ajout.

La structure du document

Un objet accessible directement depuis le pied du fichier est le catalogue du document qui agit comme racine pour la structure du document.

Ce catalogue, en forme d'objet dictionnaire, permet de trouver les composants du document.

Le composant le plus important est l'arborescence des pages qui est une structure permettant d'identifier la définition de chaque page dans le document.

Le fait que le catalogue soit un dictionnaire avec des objets nommés à l'intérieur a permis, au fur et à mesure des versions de PDF, d'ajouter d'autres structures dans le document sans casser la compatibilité avec les fichiers existants.

La description des pages

Le contenu de chaque page est décrit par un ou plusieurs objets de type flux (stream).

Le contenu de ces flux est fait d'objets PDF qui forment des instructions décrivant comment l'encre sera posée sur la page.

Exemple de fichier PDF

Voilà le texte d'un fichier PDF très simple :

```
%PDF-1.0
1 0 obj
<< /Type/Catalog /Pages 2 0 R >>
endobj
2 0 obj
<< /Type /Pages /Count 1 /Kids [3 0 R] >>
endobj
3 0 obj
<< /Type /Page /Parent 2 0 R /MediaBox [ 0 0 595.32 841.92]
/Resources << /Font << /F0 4 0 R >> >> /Contents 5 0 R >>
endobj
4 0 obj
<< /Type /Font /Subtype /Type1 /Name /F0 /BaseFont /Helvetica
/Encoding /WinAnsiEncoding >>
endobj
5 0 obj
<< /Length 205 >>
stream
BT
/F0 10 Tf
50 791 Td
(Fichier PDF simple) Tj
ET
endstream
xref
1 5
0000000010 00000 n
0000000061 00000 n
0000000121 00000 n
0000000258 00000 n
0000000368 00000 n
trailer
<< /Size 5 /Root 1 0 R >>
startxref
470
%%EOF
```

Si ce fichier est saisi dans un éditeur de texte avec des fins de ligne DOS (deux octets par fin de ligne) il peut être affiché dans un visualisateur PDF. Ça ne marcherait pas avec des fins de lignes Unix (un octet par fin de ligne) puisque plusieurs valeurs dans le fichier indiquent des positions par octet dans le fichier.

Dans cet exemple cinq objets sont définis :

- **objet 1** — c'est le catalogue du fichier qui indique simplement où trouver l'arborescence des pages (l'objet 2) ;
- **objet 2** — c'est l'arborescence des pages qui indique qu'il y a une seule page (l'objet 3) ;
- **objet 3** — définit la page comme étant de taille A4 (595,32 x 841,92 points) et qui utilise une police (l'objet 4) et dont le contenu est dans l'objet 5 ;
- **objet 4** — indique que la police appelée /F0 est Helvetica ;
- **objet 5** — définit le contenu de la page. Il appelle la police Helvetica et positionne le texte « Fichier PDF simple » en haut et à gauche de la page.

Le pied du fichier indique que cinq objets sont définis est que le catalogue du fichier est l'objet 1.

Fonctionnalités

Le format PDF est extrêmement riche. Ce qui a commencé comme un simple format de fichier pour représenter le contenu visuel des pages d'un document a évolué pour pouvoir stocker beaucoup de types de contenus numériques.

Dans cette section on décrit les fonctionnalités disponibles dans un fichier PDF, et ceci avec deux objectifs :

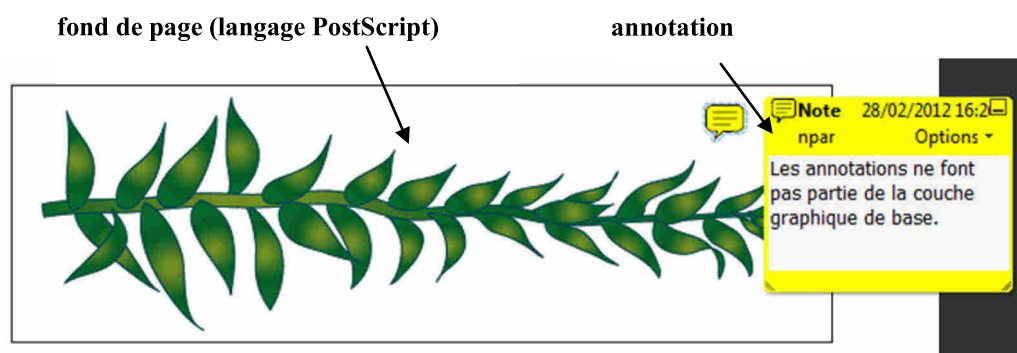
- pour ceux qui ne veulent pas aller plus loin dans l'historique des versions, cela permet de savoir ce qu'il est possible de faire avec un fichier PDF ;
- pour ceux qui sont intéressés par les détails des parties 2 et 3 de ce rapport, cela donne une liste de fonctionnalités qui sera utilisée pour catégoriser les évolutions du format de base du PDF et des normes associées.

Structure du fichier

Dans cette rubrique on mettra tout ce qui n'est pas en rapport avec les autres rubriques et qui ne concerne pas en premier lieu l'utilisateur, mais surtout le développeur de produits PDF. Ainsi les éléments qui font en sorte qu'un fichier PDF soit lisible seront décrits sous cette rubrique, mais les fonctionnalités qu'il est possible ou non d'inclure dans un fichier seront incluses dans d'autres rubriques.

Modélisation de l'image

Un fichier PDF peut être vu comme ayant deux couches. Une couche qui définit le fond de page (le contenu de la page comme décrit dans le langage PostScript) et tout le reste (annotations, structure textuelle, etc). Sous modélisation de l'image on désigne ce qui est en rapport avec la création du fond de la page, sauf certains aspects de cette modélisation qui sont suffisamment importants pour mériter une rubrique à part (couleur, polices, transparence).



Couleur

Cette rubrique appartient à la modélisation de l'image, mais puisque la gestion de couleur est très complexe et fait l'objet de beaucoup de pages dans les spécifications et normes PDF, il est traité à part. On mettra donc sous cette rubrique tout ce qui est en rapport avec la modélisation de la couleur, les espaces de couleur et le calibrage des couleurs.

Historique :

- PDF 1.0 — Gris, RVB, CMJN, indexé : dépendant du périphérique de la sortie ;
- PDF 1.1 — Gris, RVB, Lab : calibré ;
- PDF 1.2 — Séparations, motifs ;
- PDF 1.3 — Couleurs à base d'ICC, DeviceN.

Polices

La gestion des polices fait aussi partie de la modélisation de l'image, mais c'est un sujet important et il est donc traité à part.

Historique :

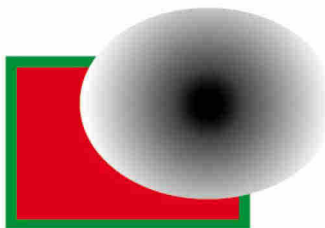
- PDF 1.0 — TrueType, Type 1 et Type 3 ;
- PDF 1.1 — jeux partiels de police ;
- PDF 1.6 — OpenType.

Transparence

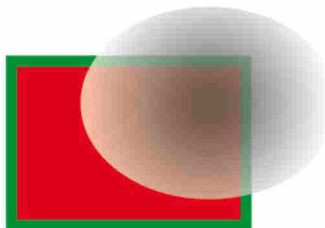
Le traitement de la transparence fait partie de la modélisation de l'image, mais puisque la transparence ne faisait pas partie de PostScript, ni des versions originales de PDF, il donne souvent lieu à des problèmes. Il est donc traité à part.

PostScript et les premières versions de PDF, cachaient toujours les objets déjà sur une page par les objets placés ultérieurement. Avec la transparence, l'objet peut plus ou moins apparaître de derrière à travers un objet placé devant.

Sans transparence :



Avec transparence :



Historique :

- PDF 1.0 — modélisation de l'image sans transparence ;
- PDF 1.4 — introduction de la modélisation avec transparence.

Impression

Sous cette rubrique on traitera tout ce qui est en rapport avec l'impression d'un document par rapport à sa visualisation à l'écran. Même si à première vue on pourrait penser que l'impression est la simple transposition de ce qu'on voit à l'écran, les technologies d'impression sont complexes, et les exigences d'un rendu papier plus strictes que pour l'écran ; nous verrons les nombreux facteurs à prendre en compte pour l'impression d'un document.

Préférences de document

C'est un ensemble de valeurs qui spécifie les paramètres d'affichage ou d'impression d'un document. Cette rubrique couvre les évolutions apportées par chaque version dans le dictionnaire des préférences d'un document, mais aussi de nouvelles préférences incluses dans un fichier PDF (comme sur la définition des pages par exemple).

Exemple de préférences de document :

Disposition et facteur de zoom

Onglet de navigation : Panneau Signets et page

Disposition des pages : Par défaut

Zoom : Par défaut

Ouvrir à la page : 1 sur 1133

Fenêtre

☐ Redimensionner selon la page initiale

☐ Centrer à l'écran

☐ Ouvrir en mode plein écran

Affichage : Nom du fichier

Historique :

- PDF 1.2 — introduction des préférences de document (nombre de pages visibles, autres fenêtres ouvertes, ...);
- PDF 1.3 — niveau de zoom par défaut pour une page ;
- PDF 1.4 — contrôle de l'affichage du titre, utilisation des régions de page ;
- PDF 1.7 — préférences d'impression (recto/verso, nombre de copies etc.) ;
- PDF 2.0 — possibilité de forcer les valeurs des préférences de document.

Métadonnées

Les métadonnées d'un document sont des données qui fournissent plus d'informations que son contenu seul : son titre, sa date de rédaction, son auteur par exemple. Des informations sur le format du fichier (son niveau de conformité par rapport à une norme par exemple) peuvent aussi y être stockées. Au début de PDF ces informations étaient contenues dans le dictionnaire des informations du document. Plus tard une façon plus souple de les stocker en XML a été introduite avec les métadonnées XMP.

Exemple de métadonnées d'un fichier :

Description

Fichier : bulletin1101.pdf

Titre : European Patent Bulletin 2011/01

Auteur : European Patent Office

Sujet : European Patent Bulletin - Europäisches Patentblatt - Bulletin européen des brevets

Mots-clés : EPO; EPA; OEB; BULLETIN; PATENTBLATT; GAZETTE; 2011/01, 201101

Date de création : 13/04/2011 16:58:38

Modifié le : 13/04/2011 17:06:52

Application : compepb © diadeis 2011

Métadonnées supplémentaires...

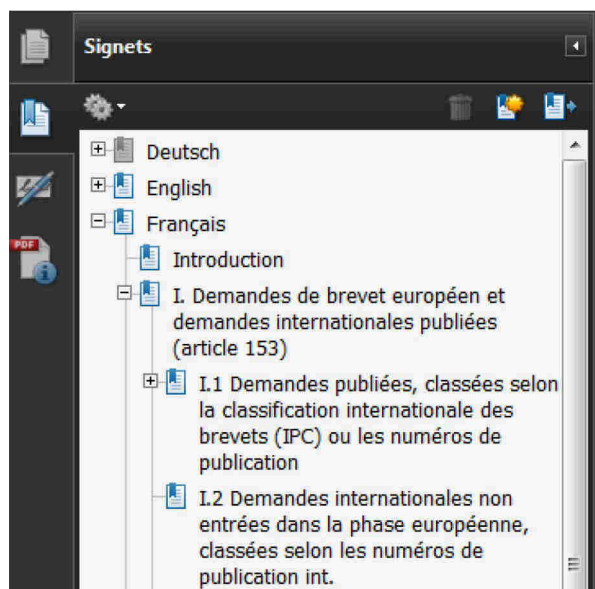
Historique :

- PDF 1.0 — dictionnaire des informations du document ;
- PDF 1.4 — métadonnées XML au format XMP.

Sommaire

Le sommaire est une liste de signets présentée comme table des matières du document, souvent dans une fenêtre à part du document dans un outil de visualisation PDF.

Exemple de sommaire de document :



Historique :

- PDF 1.0 — introduction du sommaire dans un fichier PDF.

Signatures numériques

Une signature numérique permet à une ou plusieurs personnes ou entités de signer un document. La signature peut être faite pour garantir l'authenticité du document ou pour indiquer que la personne a validé son contenu à un moment donné. L'objectif d'une signature est d'authentifier le signataire et de garantir l'état du document au moment de la signature. Le fait de signer un document peut limiter l'accès ultérieur à ce document.

Historique :

- PDF 1.3 — introduction des signatures numériques dans les fichiers PDF ;
- PDF 1.5 — signatures sur une partie de document, ajout des informations sur l'environnement lors de la signature ;
- PDF 1.6 — signature auteur, horodatage, SHA256 ;
- PDF 1.7 — SHA384, SHA512 et RIPEMD160 ;
- PDF 2.0 — CMS, CAdES, Lieu de stockage des éléments de sécurité du document, contraintes.

Articles

Un article dans PDF est un terme très précis. Il est utilisé dans le cas où un document contient plusieurs flux de texte indépendants, comme par exemple dans un magazine où on pourrait avoir un article principal et un deuxième flux de texte qui continue en encadré sur plusieurs pages.

Un article PDF identifie un de ces flux de texte, mais pas le texte lui-même, uniquement les régions (rectangulaires) sur la page qui affiche le texte de l'article. Ainsi un article est une liste ordonnée de rectangles (connu comme *beads* qu'on pourrait traduire par perles) chacune indiquant la suite de l'article sur une page.

Historique :

- PDF 1.1 — introduction des articles dans un fichier PDF.

Présentations

Une présentation est une façon de visualiser une série d'images sous la forme d'un diaporama, où il est par exemple possible de définir les transitions entre les images ou le délai d'affichage de chaque image.

Le format PDF comporte deux types de présentation :

- Mode présentation — chaque page est traitée comme une diapositive à afficher et il est possible de définir les transitions ;
- Présentations alternatives — la présentation est faite d'images stockées dans le fichier PDF (et qui apparaissent ou non sur les pages du document).

Historique :

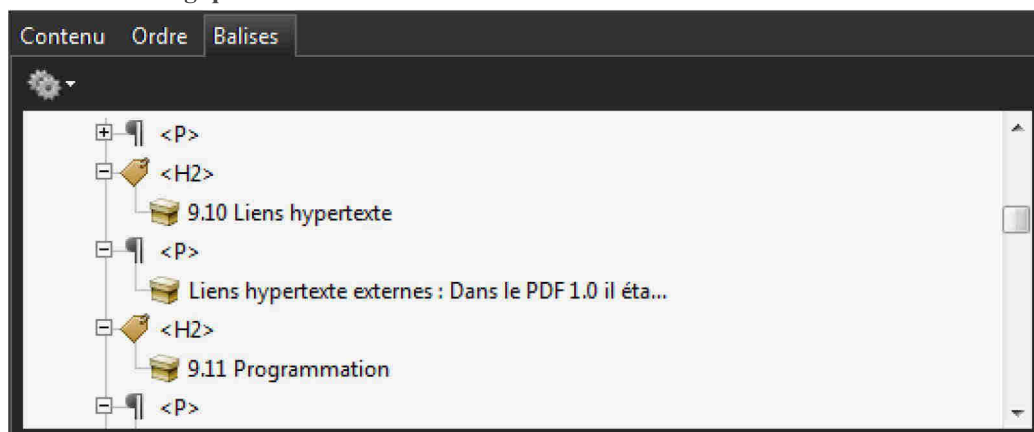
- PDF 1.1 — mode présentation ;
- PDF 1.4 — présentations alternatives ;
- PDF 1.5 — (mode présentation) l'introduction du contenu facultatif permet de créer des transitions sur des éléments contenus dans les pages, et pas uniquement sur les pages.

Structure logique de document

Le langage PostScript, et ainsi PDF à l'origine, ne définissait que les marques d'encre à mettre sur une page. Puisque ces marques étaient majoritairement des glyphes d'une police dont l'encodage était connu, il était souvent possible de récupérer le texte d'un document PDF ; mais rarement avec un taux de réussite à 100%. Par exemple, sur une page comportant deux colonnes, il n'était pas toujours possible de différencier le texte des deux colonnes, et on récupérait un mélange des deux.

La définition de la structure logique de document permet de récupérer le texte du document, mais aussi sa structure. Cette possibilité est si importante qu'elle est devenue obligatoire dans plusieurs des normes basées sur le PDF.

Exemple de structure logique dans un document :



Historique :

- PDF 1.3 — introduction de la notion de structure de document ;
- PDF 1.4 — normalisation des noms de balises (PDF balisé) ;
- PDF 1.5 — ajout de noms de balises dans la liste, attributs de présentation pour le reformatage de texte sur les petits écrans ;
- PDF 1.6 — façon d'indiquer des doutes dans la structure, indication de colonnes dans les éléments de groupe ;
- PDF 2.0 — des modifications pour rationaliser la façon de décrire la structure (travail toujours en cours lors de l'écriture de ce rapport).

Contenu facultatif

Le mécanisme de contenu facultatif (aussi appelé « couches ») permet de faire apparaître ou disparaître des contenus sur les pages du fichier PDF selon des critères préétablis. C'est particulièrement utile, par exemple, si on veut fournir plusieurs versions linguistiques d'un document dans le même fichier PDF.

Ce mécanisme permet seulement de cacher ou de faire apparaître des contenus. Il ne permet pas de reformater un document selon la présence ou l'absence d'un contenu facultatif.

Historique :

- PDF 1.5 — introduction du contenu facultatif.

Annotations

Sous cette rubrique on parlera des annotations qu'il est possible d'ajouter sur une page PDF comme des Post-It ou des flèches. Le terme « annotation » dans la documentation anglaise de PDF recouvre presque tout ce qu'il est possible d'ajouter à la page PDF, grâce à la fonctionnalité de modélisation de l'image. Dans ce document on traduit cette notion par « annotation PDF » et les autres types d'enrichissement PDF (tel que les liens hypertexte ou les clips vidéo) seront traités sous leur propres rubriques. Ici on ne parlera que de ce qu'on appelle en anglais les « markup annotations ».

Exemples d'annotations :

Cette image montre des annotations de type note, surlignage, flèche et dimension. Ces annotations ne font pas partie de la page, et peuvent ou non être visibles lors de l'impression de la page.



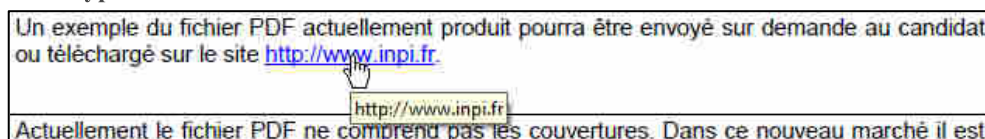
Historique :

- PDF 1.0 — les notes textuelles ;
- PDF 1.1 — amélioration dans la présentation des annotations ;
- PDF 1.3 — ajout de plus de types d'annotations (formes géométriques, tampons, mise en valeur, pièce jointe, ...) ;
- PDF 1.4 — nom de l'annotation, amélioration des caractéristiques des annotations existantes, ajout de soulignement ondulé ;
- PDF 1.5 — ajout de polygone, polyligne et caret, l'ordre de navigation des annotations, actions associées ;
- PDF 1.6 — groupement d'annotations, amélioration des caractéristiques des annotations existantes ;
- PDF 1.7 — verrouillage d'annotations, amélioration des caractéristiques des annotations existantes, annotations de type biffure ;
- PDF 2.0 — ordre d'activation des annotations.

Liens hypertexte

Sous cette rubrique on parlera des liens hypertexte à l'intérieur d'un document, entre documents PDF, ou entre un document PDF et une adresse internet quelconque.

Exemple de lien hypertexte :



Historique :

- PDF 1.0 — liens hypertexte dans un même document ;
- PDF 1.1 — liens hypertexte externe.

Compression

Sous cette rubrique on parlera de la façon dont il est possible de compresser les parties d'un fichier PDF pour en faire des fichiers plus petits. La totalité des parties d'un fichier PDF n'est pas compactable ; au fur et à mesure des versions, des possibilités ont été ajoutées pour compresser de plus en plus de parties du fichier. Les algorithmes de compression ont aussi évolué pour améliorer le taux de compression.

Historique :

- PDF 1.0 — LZW, codage par plages, CCITT Fax groupes 3 et 4, JPEG ;
- PDF 1.2 — Flate ;
- PDF 1.4 — JBIG2 ;
- PDF 1.5 — JPEG2000 et compression de plus de structures du fichier PDF.

Sécurité

Sous cette rubrique on parlera des notions de cryptage, de protection par mot de passe et des limitations du droit d'accès à un fichier PDF. Les algorithmes de cryptage disponibles et la longueur des clefs permises ont été améliorés au fur et à mesure des versions.

Historique :

- PDF 1.1 — cryptage avec mots de passe, clef de 40 bits ;
- PDF 1.3 — cryptage avec clefs publiques ;
- PDF 1.4 — clef de 128 bits ;
- PDF 1.5 — droits d'accès sans cryptage ;
- PDF 2.0 — clef de 256 bits.

Formulaires interactifs « AcroForms »

Un formulaire interactif comprend des champs qu'un utilisateur peut remplir pour pouvoir enregistrer les valeurs des champs à l'extérieur du fichier PDF. C'est utile par exemple pour créer un document de type Cerfa où le fond de page décrit le texte de la page et les champs sont disponibles pour être remplis sans que l'utilisateur puisse modifier le fond de page.

Sous cette rubrique on ne parlera que des formulaires de type « AcroForms » qui sont les formulaires interactifs de base de PDF. Le texte de base d'un formulaire « AcroForms » est le fond de page qui utilise le modèle graphique PostScript comme le reste de PDF.

Exemple d'un formulaire PDF :

Celui qui reçoit le fichier PDF peut remplir les champs, mais il ne peut pas modifier le fond du document.

I. DONATEUR(S)	
DONATEUR N° 1	
☐ Mme	☐ Mlle
☐ M	
Nom	
Prénoms	
Date de naissance	
Commune	
Département	

Historique :

- PDF 1.2 — introduction dans PDF des formulaires interactifs « AcroForms » ;
- PDF 1.4 — envoi de contenu en XML ou PDF, Unicode, améliorations de champs ;
- PDF 1.5 — champ de type peigne, verrouillage après signature, texte riche ;
- PDF 2.0 — codes-barres.

Formulaires interactifs XFA

Un deuxième type de formulaire interactif s'appelle le formulaire interactif XFA (XML Forms Architecture). Bien que le but soit identique aux formulaires de type « AcroForms », la technologie est complètement différente et il n'y a aucune interaction entre les deux types de formulaires.

XFA est un format de description de formulaire basé sur XML (XHTML 1.0 et CSS2). Il a été développé par *JetForm* et proposé comme norme au W3C. *Adobe Systems* a racheté *JetForm* en 2002.

Les formulaires XFA sont dynamiques. Le texte de base est décrit en XML et peut bouger avec les modifications dynamiques des champs du formulaire. Il devient possible par exemple de changer la taille d'un champ (et la position de tout ce qui suit) selon le contenu du champ, ce qui n'est pas possible avec les formulaires PDF. Le nombre de pages dans le fichier PDF peut même varier avec l'utilisation de formulaires interactifs XFA.

Les formulaires de type « AcroForms » sont mieux intégrés dans PDF que les formulaires XFA, mais ils sont moins souples. Les formulaires XFA sont plus dynamiques mais mal intégrés dans le PDF dans son ensemble.

Une confrontation complète (en anglais) des deux formats peut être trouvée à l'adresse : <http://acrobatusers.com/blogs/thomp/so-what-difference-between-acroforms-and-xfa>.

Historique :

- PDF 1.5 — introduction dans PDF des formulaires interactifs XFA ;
- PDF 1.7 — régénération de l'affichage lors de l'ouverture de document ;
- PDF 2.0 — enregistrement des données du formulaire dans dictionnaire, utilisation Flash.

Multimédia

Sous cette rubrique on parlera des différentes façons possibles d'intégrer l'audio et vidéo dans un fichier PDF.

Historique :

- PDF 1.2 — utilisation d'audio embarqué ou externe et vidéo externe ;
- PDF 1.5 — possibilité d'embarquer audio et vidéo.

Illustrations 3D

À l'origine PDF permettait de visualiser des pages en deux dimensions. Des extensions ultérieures ont permis la visualisation d'éléments en trois dimensions. Sous cette rubrique on parlera des différentes façons de le faire.

Historique :

- PDF 1.6 — modélisation 3D avec U3D ;
- PDF 2.0 — modélisation 3D avec PRC.

Programmation

Sous cette rubrique on parlera des différentes façons d'ajouter des actions dans un fichier PDF par des éléments PDF prédéfinis ou par des programmes JavaScript.

Historique :

- PDF 1.1 — actions prédéfinies ;
- PDF 1.3 — JavaScript.

Fichiers embarqués

Sous cette rubrique on parlera des façons d'embarquer dans un fichier PDF des fichiers qui sont ou non liés au contenu du fichier PDF, de façon à pouvoir les avoir toujours à disposition avec le fichier PDF.

Historique :

- PDF 1.3 — fichiers embarqués ;
- PDF 1.4 — liste de fichiers embarqués ;
- PDF 1.7 — collections de fichiers embarqués ;

Guide Méthodologique

Le format PDF

- PDF 2.0 — associations entre contenu PDF et fichiers embarqués.

OPI

OPI (*Open Prepress Interface*) est un mécanisme qui permet de mettre des images de basse résolution dans un fichier pour l'alléger lors de sa création, pour les remplacer par des images haute résolution lors de son impression finale. Sous cette rubrique on parlera des mécanismes mis en place dans un fichier PDF pour implémenter OPI.

Historique :

- PDF 1.2 — introduction du mécanisme d'OPI.

CAO

Sous cette rubrique on parlera des fonctionnalités PDF qui ont été mises en place spécifiquement pour aider dans la création de fichiers PDF dans le domaine de la conception assisté par ordinateur (CAO).

Historique :

- PDF 1.6 — fonctions de mesure et propriétés d'éléments.

Informations géomatiques

Sous cette rubrique on parlera des façons de lier des informations géomatiques de localisation à des éléments d'un fichier PDF.

Historique :

- introduit dans PDF 2.0.

Capture de pages web en PDF

Sous cette rubrique on parlera des fonctionnalités PDF qui ont été mises en place pour faciliter la capture de pages web en PDF et de garder les liens entre le contenu capturé et sa source.

Historique :

- PDF 1.3 — introduction des fonctionnalités pour stocker les informations de capture de page web.

Tableau récapitulatif des fonctionnalités

Ce tableau indique de façon globale l'introduction de chaque fonctionnalité potentiellement présente dans un fichier PDF, et les versions qui ont amélioré la fonctionnalité de façon significative. Le but du tableau indique la version d'introduction de la fonctionnalité, et si celle-ci a été améliorée depuis.

Légende :

	Aucune fonctionnalité
	Fonctionnalité présente
	Amélioration significative dans la fonctionnalité

On a placé les abréviations des principales améliorations dans les cellules du tableau. Le lecteur est invité à consulter le chapitre précédent pour avoir une liste plus détaillée des améliorations.

Fonctionnalités	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	2.0
Couleur		Cal	Sép	ICC					
Polices							OT		
Transparence									
Préférences de document									
Métadonnées	Dict				XMP				
Sommaire									
Signatures numériques									
Articles									
Présentations					Alt.				
Structure logique de document									
Contenu facultatif									
Annotations									
Liens hypertexte	Int	Ext							
Compression			Flate		JB2	JP2			
Sécurité		40			128				256
Formulaires interactifs « AcroForms »									
Formulaires interactifs XFA									
Multimédia			Aud			Vid			
Illustrations 3D							U3D		PRC
Programmation				JS					
Fichiers embarqués									
OPI									
CAO									
Informations géomatiques									
Capture de pages web en PDF									
Abréviations Alt =alternatives, Aud =audio, Cal =calibré, Dict =dictionnaire, Ext =externe, Int =interne, JB2 =JBIG2, JP2 =JPEG2000, JS =JavaScript, OT =OpenType, Sép =séparations, Vid =vidéo									

Introduction aux normes basées sur le PDF

Jusque-là on a parlé des versions de PDF. Le format PDF étant très riche, il a vite été constaté qu'il fallait limiter ce qu'on mettait dans un fichier PDF en fonction de son utilisation. Par exemple, il est clair que si on met une vidéo dans un fichier, le rendu qu'on aura sur une imprimante n'est pas très prévisible.

Les premiers à définir une normalisation du format PDF ont été les métiers de l'imprimerie qui ont voulu, les premiers, spécifier comment créer un PDF qui serait imprimable sans ambiguïté. Ce travail, appelé PDF/X, a commencé aux États-Unis avec une norme en 1999. Deux ans plus tard cette norme a été adaptée par l'ISO pour en créer une norme internationale.

D'autres métiers ont aussi constaté l'utilité de PDF et ont commencé à produire d'autres normes. Pour la plupart ces normes limitent ce qu'il est possible de mettre dans un PDF, par rapport à un fichier PDF de base, mais parfois ils proposent aussi des ajouts dans le format de base, qui sont, la plupart de temps, repris dans le format de base par la suite.

Il y a deux comités chez l'ISO qui travaillent sur le PDF :

- TC 130 pour tout ce qui est en rapport avec les métiers de l'imprimerie et les arts graphiques ;
- TC 171 pour tout ce qui est en rapport avec les documents électroniques.

Voilà une liste pour présenter chacune des normes ISO :

PDF (ISO 32000)

Depuis le PDF 1.7, le PDF est contrôlé par l'ISO au lieu d'Adobe. Les spécifications du format de base sont donc publiées sous la norme ISO-32000.

PDF/A (ISO 19005)

Des normes pour l'archivage des documents électroniques.

PDF/E (ISO 24517)

Des normes pour l'utilisation de PDF dans les métiers de l'ingénierie.

PDF/X (ISO 15930)

Des normes pour l'utilisation de PDF dans l'imprimerie et les métiers d'arts graphiques.

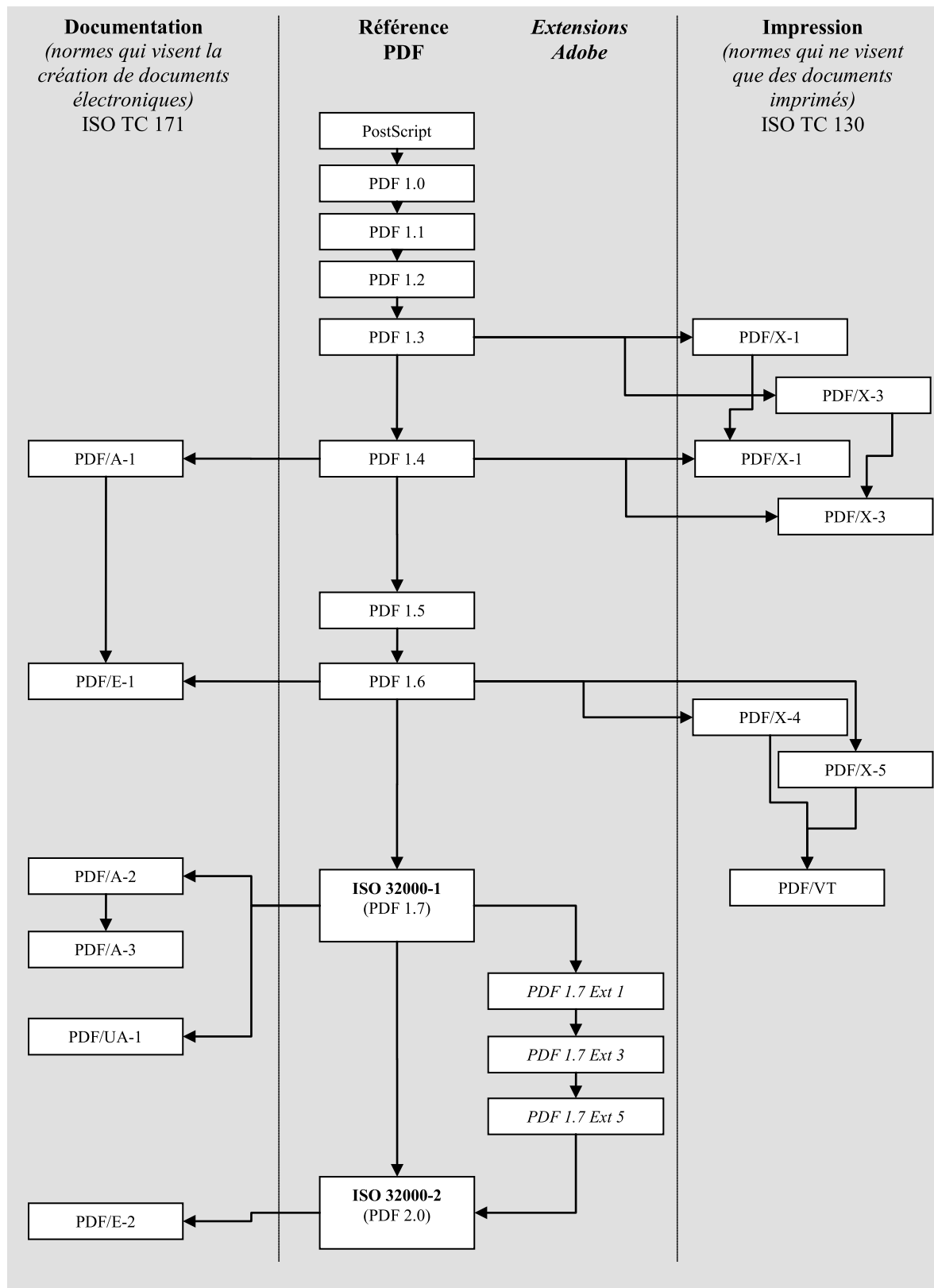
PDF/UA (ISO 14289)

Des normes pour indiquer comment créer des documents PDF qui seront accessibles au plus grand nombre de personnes (en tenant compte des handicaps possibles).

PDF/VT (ISO 16612)

Des normes sur l'utilisation de PDF dans les métiers de l'édition.

Schéma des liens entre les versions de PDF



Adobe Acrobat

Bien que le format de fichier PDF ne soit pas obligatoirement lié au logiciel Acrobat, le développement du logiciel a toujours été le moteur des améliorations dans le format. Ce n'est que depuis le transfert du contrôle de PDF vers l'ISO que les versions de PDF ne sont plus liées aux versions d'Acrobat.

C'est pour ces raisons qu'on inclut ici un tableau pour indiquer les correspondances entre les versions de PDF et les versions d'Acrobat, indiquant aussi les changements dans le produit Adobe.

Version Acrobat	Version PDF	
1.0 (juin 1993)	1.0	Cette version d'Acrobat comporte trois composants : <i>Acrobat Exchange</i> — manipulation de PDF <i>Acrobat Distiller</i> — création de PDF depuis PostScript <i>Acrobat Reader</i> — lecture des fichiers PDF
2.0 (septembre 1994)	1.1	<i>Acrobat Exchange</i> reçoit une architecture avec support pour des plug-ins et la recherche dans les fichiers PDF
2.1 (août 1995)		Support pour audio et vidéo.
3.0 (novembre 1996)	1.2	Visualisation de PDF dans les navigateurs web, utilisation de formulaires interactifs.
4.0 (avril 1999)	1.3	<i>Acrobat Exchange</i> est renommé <i>Acrobat</i> .
	1.4	PDF 1.4 fait son apparition dans <i>Illustrator 9</i> (mi-2000) — utilisation de la transparence.
5.0 (mai 2001)		Améliorations dans l'utilisation des annotations, affichage de surimpression, utilisation des formulaires interactifs. Meilleure intégration dans <i>Microsoft Office</i> .
6.0 (juillet 2003)	1.5	<i>Acrobat Reader</i> est renommé <i>Adobe Reader</i> et permet de lire des fichiers eBook. Création de <i>Acrobat Professional</i> pour l'utilisation prépresse.
7.0 (janvier 2005)	1.6	Affichage des objets 3D.
8.0 (novembre 2006)	1.7	Acrobat 8 continue de proposer PDF 1.6 comme format d'enregistrement par défaut.
8.1 (juin 2007)	1.7 niveau d'extension 1	Support Office 2007, Vista et Windows 64-bit.
9.0 (juin 2008)	1.7 niveau d'extension 3	Comparaison de documents PDF, vidéo en format Flash, cartes géomatiques.
9.1 (juillet 2009)	1.7 niveau d'extension 5	Mise à jour de sécurité.
X (novembre 2010)		Simplification de l'interface utilisateur. Aucune publication d'une nouvelle version de PDF.

Conclusions

Le format PDF, introduit en 1993 a bien fait ses preuves. Il existe peu de formats de fichier qui ont réussi à évoluer autant que le PDF tout en gardant une compatibilité ascendante et descendante entre les différentes versions du fichier.

Guide Méthodologique

Le format PDF

Après presque 20 ans, il est toujours possible de lire les fichiers PDF 1.0 avec les lecteurs d'aujourd'hui, et en principe, il serait possible de lire au moins le contenu de la page d'un PDF actuel, avec un lecteur original. Il est donc possible de conclure que le format est à la fois solide et évolutif. Le fait que ce format ait été utilisé comme norme pour l'archivage n'est pas le fruit du hasard — avec un tel historique, il présente de solides gages de pérennité pour l'avenir.

Comment cela va-t-il évoluer ?

Il est probable que la vitesse d'évolution va ralentir. Maintenant que le format est contrôlé par l'ISO, au lieu d'une entreprise, les changements seront pilotés par les demandes des différents groupes d'utilisateurs au lieu d'être exigés par des besoins d'avoir des nouveautés à vendre.

Déjà, on ressent que les fonctionnalités qui s'insèrent mal dans le format, comme les formulaires interactifs XFA, ne seront pas mis en avant et même peut-être relégués aux oubliettes. Avec la sortie prochaine de la version 2.0, il sera peut-être temps de faire une pause dans la version de base pour laisser les normes associées rattraper leur retard. La multiplicité de versions et le fait que chaque norme se base sur une version différente, n'aide guère dans la compréhension.

PDF 2.0 sera la première version du format créée par l'ISO. Pour converger vers une famille de normes cohérentes, il faudra que ces normes soient mises à jour pour utiliser le format PDF 2.0.

La prochaine étape sera probablement la simplification. PDF est devenu trop complexe. Il y a trop de possibilités et trop de manières de faire une seule et même chose. Indiquer quelles sont les fonctions de base que tout lecteur doit supporter, et quelles sont les fonctions spécialisées, utiles pour des marchés particuliers, devient indispensable.

Ce qui est certain, c'est que le format a un bel avenir. Il est incontournable et risque de l'être pendant un bon moment encore.

Le monde informatique est jeune et change très vite. Personne ne sait comment seront nos ordinateurs dans 20 ans. Mais ce qui semble assez certain, c'est qu'ils seront toujours capables de lire des fichiers PDF.

Glossaire général

AFP (*Advanced Function Presentation*) — un langage développé par IBM et utilisé dans l'édition (<http://www.outputlinks.com/SpecialInterest/AFPCColorConsortium/home.html>).

CIE (*Commission internationale de l'éclairage*) — les espaces de couleur définis par la CIE permettent de décrire les couleurs en relation avec la perception visuelle humaine (<http://cie.co.at/>).

CMJN (Cyan/Magenta/Jaune/Noir) — les couleurs de base utilisées pour l'impression en synthèse soustractive.

CMYK (Cyan/Magenta/Yellow/Black) — les couleurs de base utilisées pour l'impression en synthèse soustractive (acronyme anglais).

DSC (*Document Structuring Conventions*) — une version d'EPS qui permet de définir des séparations de couleur.

ECMA (originellement *European Computer Manufacturers Association*) — organisation internationale de normalisation technologique (<http://www.ecma-international.org/>).

EPS (*Encapsulated PostScript*) — un format d'image vectorielle basé sur le PostScript.

n-colorant — modèle de couleurs autre que gris, RVB ou CMJN qui comporte au moins deux colorants

OPI (*Open Prepress Interface*) — un système de remplacement d'images permettant une manipulation aisée de fichiers de publication assistée par ordinateur (PAO) grâce à l'allègement du poids des fichiers.

PPML (*Personalized Print Markup Language*) — un langage basé sur XML utilisé dans l'édition (<http://ppml.podi.org/>).

PRC (*Product Representation Compact*) — un format de fichiers pour décrire des objets 3D. Développé par *Trade and Technologies France* (TTF) qui a été racheté par Adobe en 2006.

PRONOM — un registre technique, utilisant une technologie web, développé par les archives nationales du Royaume Uni qui sert à identifier des formats de fichier (<http://www.nationalarchives.gov.uk/PRONOM/>).

RFC (*Request for Comments*) — des documents de spécifications liés aux technologies du web et publiés par l'IETF (*Internet Engineering Task Force*) (<http://www.ietf.org/rfc.html>).

RGB (Red/Green/Blue) — les couleurs de base utilisées pour la visualisation écran en synthèse additive (acronyme anglais).

RVB (Rouge/Vert/Bleu) — les couleurs de base utilisées pour la visualisation écran en synthèse additive.

SWF — un format de fichiers pour le multimédia, les graphiques vectoriels et des scripts dans l'environnement *Flash* d'Adobe. Quand on parle du format SWF ce sont des éléments multimédia comme on verrait dans un navigateur Web avec le plug-in *Flash*.

TIFF (*Tagged Image File Format*) — un format d'images rasterisées défini par Aldus et maintenant contrôlé par Adobe. Il n'a pas eu de révision majeure depuis 1992.

TIFF/IT (*TIFF/Image Technology*) — une norme (ISO 12639) utilisée par l'industrie du prépresse pour décrire des pages par des images rasterisées, basée sur le TIFF.

U3D (*Universal 3D file format*) — un format de fichiers pour décrire des objets 3D, développé par le *3D Industry Forum*. Ce format a été normalisé par l'ECMA comme ECMA-363 (<http://www.ecma-international.org/publications/standards/Ecma-363.htm>).

Unicode — Un codage de caractères qui se veut universel (<http://www.unicode.org/>). Équivalent à l'ISO/CEI 10646.

Glossaire PDF

La quasi-totalité des documents et normes concernant PDF sont en anglais. Lors de l'écriture de ce rapport on a dû traduire les termes techniques utilisés de l'anglais en français, tout en essayant de rester cohérent dans les termes utilisés. Ci-dessous, une liste des traductions pour les termes techniques du monde PDF :

Français vers anglais

Français	Anglais	Sens
annotation PDF	annotation	Un type de contenu PDF qui peut être une annotation ajoutée à la page, un élément multimédia ou un lien hypertexte.
annotation	markup annotation	Un élément ajouté à la page pour apporter des précisions ou indiquer des corrections.
annotation de type biffure	redaction annotation	Une annotation qui indique du contenu à supprimer.
besoins du document	document requirements	Une liste des fonctionnalités nécessaires à la lecture correcte d'un fichier PDF.
besoins du lecteur	reader requirements	Une liste des exigences, dans la façon de traiter son contenu, qu'un fichier PDF peut imposer sur un lecteur.
collection de fichiers embarqués	portable collection	Une façon de présenter à l'utilisateur une collection de fichiers embarqués.
dictionnaire des informations du document	document information dictionary	Un dictionnaire qui stocke les métadonnées pour un document. Avant l'introduction des métadonnées XMP c'était le seul lieu de stockage des métadonnées.
fichier associé	associated file	Un fichier embarqué dont le lien avec le fichier PDF est explicite.
fichier embarqué	embedded file	Tout fichier inclus dans un fichier PDF qui ne fait pas partie des fonctionnalités PDF du document.
flux	stream	Un des types d'objets de base d'un fichier PDF. Permet de stocker n'importe quel série d'octets, mais souvent le contenu est une suite des autres types d'objets PDF.
fonction de transfert	transfer function	Fonction permettant de préciser une façon non-linéaire de rendre les différents niveaux de couleur.
forme composite	form XObject	Forme créée avec des opérateurs de dessin mais mise dans un objet pour pouvoir être appliquée plusieurs fois dans un fichier.
gestionnaire des besoins	requirement handler	Méthode JavaScript pour vérifier les besoins d'un document.
horodatage au niveau document	document level time stamp	Preuve du contenu du fichier à un moment précis.
intention de sortie	output intent	Des indications qui permettent de faire correspondre des caractéristiques de couleur décrites dans un fichier PDF avec celles d'un périphérique de sortie cible.

Guide Méthodologique

Le format PDF

jeu partiel de police	font subset	Façon d'embarquer une police dans un fichier PDF en incluant uniquement les caractères nécessaires au lieu d'inclure la totalité des caractères de la police d'origine.
lieu de stockage des éléments de sécurité du document	document security store (DSS)	Un lieu pour stocker les éléments de sécurité liés au document.
modélisation de l'image	imaging model	Façon dont on décrit l'image de fond d'une page PDF.
motif	pattern	Un espace de couleurs utilisé pour remplir des régions avec un motif répété.
navigateur	navigator	Visualisation d'une collection de fichiers embarqués.
partie de document (DPart)	document part (DPart)	Une page ou une séquence de pages identifiées comme étant une unité de traitement, surtout pour l'édition.
PDF Balisé	Tagged PDF	Un système de balises normalisées utilisé pour décrire la structure logique d'un document.
PDF imbriqué	reference XObject	Un mécanisme pour inclure un fichier PDF comme image dans un autre fichier PDF.
PDF linéarisé	linearized PDF	Une façon de structurer un fichier PDF pour l'optimiser pour un affichage web.
préférences de document	viewer preferences	Un ensemble de valeurs qui exprime les préférences d'un document par rapport à l'affichage ou à l'impression.
présentation alternative	alternate presentation	Un diaporama fait d'images qui sont stockées dans le fichier PDF (et qui apparaissent ou non sur les pages du document).
signature numérique	digital signature	Élément ajouté à un fichier PDF, dont la valeur est calculée selon le contenu du fichier et d'une clef unique qui permet d'identifier le signataire et de valider l'état du document au moment de la pose de la signature.
signet	bookmark	Un élément du sommaire.
sommaire	document outline	La liste des signets, présentée comme table des matières du document.
vignette	thumbnail	Image miniature pour représenter une page ou un fichier.

Anglais vers français

Anglais	Français	Sens
alternate presentation	présentation alternative	Un diaporama fait d'images qui sont stockées dans le fichier PDF (et qui apparaissent ou non sur les pages du document).
annotation	annotation PDF	Un type de contenu PDF qui peut être une annotation ajoutée à la page, un élément multimédia ou un lien hypertexte.
associated file	fichier associé	Un fichier embarqué dont le lien avec le fichier PDF est explicite.
bookmark	signet	Un élément du sommaire.
digital signature	signature numérique	Élément ajouté à un fichier PDF, dont la valeur est calculée selon le contenu du fichier et d'une clef unique qui permet d'identifier le signataire et de valider l'état du document au moment de la pose de la signature.
document information dictionary	dictionnaire des informations du document	Un dictionnaire qui stocke les métadonnées pour un document. Avant l'introduction des métadonnées XMP c'était le seul lieu de stockage des métadonnées.
document level time stamp	horodatage au niveau document	Preuve du contenu du fichier à un moment précis.
document outline	sommaire	La liste des signets, présentée comme table des matières du document.
document part (DPart)	partie de document (DPart)	Une page ou une séquence de pages identifiées comme étant une unité de traitement, surtout pour l'éditique.
document requirements	besoins du document	Une liste des fonctionnalités nécessaires à la lecture correcte d'un fichier PDF.
document security store (DSS)	lieu de stockage des éléments de sécurité du document	Un lieu pour stocker les éléments de sécurité liés au document.
embedded file	fichier embarqué	Tout fichier inclus dans un fichier PDF qui ne fait pas partie des fonctionnalités PDF du document.
font subset	jeu partiel de police	Façon d'embarquer une police dans un fichier PDF en incluant uniquement les caractères nécessaires au lieu d'inclure la totalité des caractères de la police d'origine.
form XObject	forme composite	Forme créée avec des opérateurs de dessin mais mise dans un objet pour pouvoir être appliquée plusieurs fois dans un fichier.
imaging model	modélisation de l'image	Façon dont on décrit l'image de fond d'une page PDF.
linearized PDF	PDF linéarisé	Une façon de structurer un fichier PDF pour l'optimiser pour un affichage web.
markup annotation	annotation	Un élément ajouté à la page pour apporter des précisions ou indiquer des corrections.
navigator	navigateur	Visualisation d'une collection de fichiers embarqués.

Guide Méthodologique

Le format PDF

output intent	intention de sortie	Des indications qui permettent de faire correspondre des caractéristiques de couleur décrites dans un fichier PDF avec celles d'un périphérique de sortie cible.
pattern	motif	Un espace de couleurs utilisé pour remplir des régions avec un motif répété.
portable collection	collection de fichiers embarqués	Une façon de présenter à l'utilisateur une collection de fichiers embarqués.
reader requirements	besoins du lecteur	Une liste des exigences, dans la façon de traiter son contenu, qu'un fichier PDF peut imposer sur un lecteur.
redaction annotation	annotation de type biffure	Une annotation qui indique du contenu à supprimer.
reference XObject	PDF imbriqué	Un mécanisme pour inclure un fichier PDF comme image dans un autre fichier PDF.
requirement handler	gestionnaire des besoins	Méthode JavaScript pour vérifier les besoins d'un document.
stream	flux	Un des types d'objets de base d'un fichier PDF. Permet de stocker n'importe quel série d'octets, mais souvent le contenu est une suite des autres types d'objets PDF.
Tagged PDF	PDF Balisé	Un système de balises normalisées utilisé pour décrire la structure logique d'un document.
thumbnail	vignette	Image miniature pour représenter une page ou un fichier.
transfer function	fonction de transfert	Fonction permettant de préciser une façon non-linéaire de rendre les différents niveaux de couleur.
viewer preferences	préférences de document	Un ensemble de valeurs qui exprime les préférences d'un document par rapport à l'affichage ou à l'impression.

Bibliographie

Ne sont pas mentionnés ici les documents (spécifications PDF et normes ISO) qui font l'objet de chapitres entiers dans les parties 2 et 3 de cette étude.

- *Manuel de référence du langage PostScript (deuxième édition)*, Adobe Systems Incorporated, Éditions Addison-Wesley France, 1992. ISBN 2-87908-009-6.
- *PostScript Language Reference Manual*, Adobe Systems Incorporated, Addison-Wesley Publishing Company, 1985. ISBN 0-201-10174-2.
- *PostScript Language Reference Manual (second edition)*, Adobe Systems Incorporated, Addison-Wesley Publishing Company, 1990. ISBN 0-201-18127-4.
- *PostScript Language Reference Manual (third edition)*, Adobe Systems Incorporated, Addison-Wesley Publishing Company, 1999. ISBN 0-201-37922-8.
- *The Camelot Project*, John Warnock, 1991 (memo interne Adobe)

Les versions de PDF

PDF v1.0

Version PDF	1.0
Version Acrobat	1.0
Date de sortie	15 juin 1993
Identifiant PRONOM	fnt/14

La première version de PDF (1.0) correspondait à la version 1.0 du logiciel Adobe Acrobat. Ce logiciel a été vendu à partir du 15 juin 1993 et le format PDF a été tout de suite publié dans un livre édité par Addison Wesley. Au début ce logiciel était disponible uniquement sur Macintosh, mais une version PC a été ajoutée plus tard.

La première version de PDF comportait les fonctionnalités suivantes :

Structure du document

Arborescence de pages (pages tree) : Depuis le catalogue du document, il est possible d'accéder directement à la racine de l'arborescence de pages. C'est une structure qui peut être linéaire ou en forme d'arbre qui permet au logiciel qui écrit le fichier PDF de générer un accès rapide à chaque page. Plus le document est grand, plus il est important de générer l'arbre de façon à limiter le nombre d'accès nécessaires pour trouver une page.

Description des pages : Chaque page est décrite avec des instructions qui correspondent aux instructions PostScript de pose d'encre sur la page. Il est aussi possible d'associer une icône miniature pour représenter la page dans un sommaire graphique

ProcSet : Il est obligatoire dans la version 1.0 de signaler, avec l'aide d'un ProcSet, la liste des types d'instruction utilisés dans la description d'une page. C'est une liste des types d'instruction utilisés dans la description d'une page. Elle est nécessaire uniquement pour indiquer les définitions de fonctions à envoyer à une imprimante lors de l'impression PostScript.

Forme composite : Il est possible de définir une séquence d'instructions pour définir une forme qui peut être incluse à plusieurs endroits sur une page ou sur plusieurs pages.

Couleur

PDF 1.0 définit plusieurs espaces de couleur :

- Gris (DeviceGray) — couleur définie par des niveaux de gris ;
- RVB (DeviceRGB) — couleur définie par des composants rouge, vert et bleu ;
- CMJN (DeviceCMYK) — couleur définie par des composants cyan, magenta, jaune et noir ;
- Indexé — pour référencer une collection de couleurs par un numéro d'index. Chaque couleur est définie par une couleur RVB ou CMJN. Ce n'est pas la même chose que les couleurs de séparation introduites avec PDF 1.2.

Polices

Pour la gestion des polices, PDF 1.0 définit plusieurs types de ressources :

- Police — PDF 1.0 permet l'inclusion de polices Type1, Type 1 Multiple Master, Type 3 et TrueType. 14 polices sont identifiées comme étant toujours présentes et donc il n'est pas nécessaire de les inclure dans un document PDF pour qu'elles soient affichées :
 - Courier ;

Guide Méthodologique

Le format PDF

- Courier-Bold ;
- Courier-Oblique ;
- Courier-BoldOblique ;
- Times-Roman ;
- Times-Bold ;
- Times-Italic ;
- Times-BoldItalic ;
- Helvetica ;
- Helvetica-Bold ;
- Helvetica-Oblique ;
- Helvetica-BoldOblique ;
- ZapfDingbats ;
- Symbol.
- Encodage — décrit la correspondance entre les caractères inclus dans les chaînes de caractères d'une page et les glyphes d'une police ;
- Description de police (FontDescriptor) — définit les chasses, les attributs et les glyphes d'une police pour permettre sa substitution si la police originale n'est pas accessible.

Les 14 polices de base ont été choisies pour pouvoir servir comme polices de substitution si d'autres polices manquaient. Ainsi on avait quatre variations (romain, gras, italique et gras-italique) des trois types de police de base (police à chasse fixe, polices avec et sans empattement) plus deux polices de symboles.

Impression

Chaque page peut avoir deux dimensions :

- **MediaBox** — la taille du support sur lequel on imprimera la page ;
- **CropBox** — le rectangle à l'extérieur duquel on ne met pas de marques sur la page.

Il est ainsi possible de définir une région de fond perdu.

En plus du contenu de la page, une rotation (en multiple de 90°) peut être associée à chaque page.

Métadonnées

Un fichier PDF comporte un dictionnaire des informations du document qui permet de préciser :

- l'auteur ;
- la date de création ;
- le logiciel utilisé pour créer le document dans son format originel ;
- le logiciel utilisé pour convertir le document en PDF depuis son format originel.

Sommaire

Un sommaire est représenté dans un fichier PDF par une structure hiérarchique de signets qui indiquent chacun le texte à faire apparaître dans le sommaire et la page (et partie de la page) cible si on clique sur le signet.

Annotations

PDF 1.0 définit un type d'annotation : des notes textuelles.

Liens hypertexte

PDF 1.0 permet de définir des liens hypertexte, mais uniquement vers d'autres pages du même document.

Compression

Dans la version 1.0 de PDF uniquement les flux (streams) peuvent être compressés.

Les algorithmes permis sont les suivants :

- LZW ;
- Codage par plages (run length encoding) ;
- CCITT Fax Groupe 3 ou Groupe 4 ;
- Transformée en cosinus discrète (Discrete cosine transform) basé sur la technique de compression des images JPEG.

Évidemment seul le premier permet de compresser les flux génériques, les autres servent à la compression d'images rasterisées dans le fichier.

Puisque la version 1.0 de PDF exige que le fichier ne contienne que des caractères 7-bit, il est nécessaire d'appliquer un codage ASCII après ces compressions pour respecter cette exigence.

PDF v1.1

Version PDF	1.1
Version Acrobat	2.0
Date de sortie	Acrobat : septembre 1994, spécification PDF : mars 1996
Identifiant PRONOM	fnt/15

Le document de référence de PDF 1.1 n'est plus disponible. Cette description est basée sur ce qu'il est possible de déduire des versions ultérieures de la spécification.

Structure du fichier

Fichiers binaires : La version 1.0 définissait un fichier PDF comme utilisant toujours des caractères 7-bit. La version 1.1 a enlevé cette restriction pour permettre des caractères 8-bit dans les fichiers. Il devient ainsi possible de stocker des informations binaires directement dans le fichier sans recours à un codage ASCII. Cette modification a pour effet de réduire la taille des fichiers.

Un des problèmes trouvés avec l'utilisation de fichiers uniquement 7-bit était que des logiciels (comme encore FTP aujourd'hui) pouvaient décider qu'un tel fichier était un fichier « texte » et modifier les fins de lignes trouvées dans le fichier et ainsi rendre le fichier PDF inutilisable. Depuis la version 1.1, il a été recommandé d'inclure systématiquement des caractères 8-bit dans un fichier PDF pour être sûr qu'il soit traité en tant que fichier binaire.

Format normalisé de dates : PDF 1.1 définit un format de chaîne de caractères utilisé pour stocker une date de façon normalisée. Il devient ainsi possible de comparer par programme les dates stockées dans un fichier PDF.

Syntaxe : PDF 1.1 exige que tout flux (stream) soit un objet indirect. Ce n'était pas une exigence de PDF 1.0, et il était possible de créer des fichiers PDF dont l'interprétation des données était ambiguë.

Identifiant unique de fichier : PDF 1.1 a introduit la notion de liens hypertexte entre un fichier PDF et un autre, mais le problème reste de savoir comment trouver le fichier cible. Vu les différentes façons de nommer un fichier sur des différents systèmes d'exploitation et le fait que le fichier cible peut être renommé, déplacé ou modifié, une notion d'identifiant a été introduite pour identifier chaque fichier PDF de façon unique. Ainsi il devient possible de vérifier qu'un fichier trouvé comme cible lors de la lecture est bien celui qui existait lors de la création du lien hypertexte.

Un identifiant de fichier PDF est donc défini qui est censé être unique pour chaque fichier. Il comporte deux parties — une partie créée lors de la création originelle du fichier et qui ne change pas, et une partie qui change avec chaque mise à jour du fichier. Un lecteur peut donc vérifier l'exactitude du fichier cible et voir s'il a été modifié.

Une suggestion sur la façon de créer l'identifiant est fournie, mais ce n'est pas un algorithme exact. Il n'est pas nécessaire de pouvoir reproduire la création de l'identifiant, mais simplement de créer une valeur qui a *des chances* d'être unique.

Modélisation de l'image

Opérateur de compatibilité : Vu le fait que PDF 1.1 définit des opérateurs qui n'étaient pas connus dans le 1.0, la spécification décrit une méthode pour entourer les éléments nouveaux pour qu'un lecteur ancien puisse les ignorer de façon à continuer sa lecture du fichier sans être perturbé.

Inclusion de code PostScript : PDF 1.1 définit des façons d'inclure du code PostScript directement dans un fichier PDF pour être interprété lors de son impression. Évidemment ça a un sens uniquement si le PDF est imprimé sur une imprimante PostScript, mais à l'époque c'était le but principal d'un fichier PDF.

Un avertissement clair indique que cette fonctionnalité ne doit être utilisée que si aucune autre façon d'arriver au même résultat n'existe dans PDF.

Couleur

Couleur indépendante du périphérique de sortie : Dans PDF 1.0 les couleurs définies étaient toujours dépendantes du périphérique de sortie. Une couleur était définie uniquement par les pourcentages des encres qui la composent, sans tenir compte des différences de rendu sur des périphériques de sortie.

PDF 1.1 définit des espaces de couleur indépendants du périphérique de sortie :

- CalGray ;
- CalRGB ;
- Lab.

Ces espaces de couleur se basent sur des correspondances avec un espace de couleurs normalisé par la CIE (*Commission Internationale de l'Éclairage*) : si un périphérique de sortie est capable de les reproduire, les couleurs seront normalisées et plus proches de ce qui est attendu.

Note : PDF 1.1 définit partiellement un espace de couleurs CalCMYK pour créer un espace de couleurs CMJN calibré. Cette définition n'a pas été retenue en PDF 1.2 et peut être considérée comme à ignorer.

Polices

Jeux partiels de police : Avec la version 1.0 de PDF, si on voulait inclure une police dans un fichier PDF il fallait inclure toute la police. À partir de 1.1 il est possible d'inclure des jeux partiels de polices pour les polices Type 1 et TrueType. L'avantage d'un jeu partiel est qu'il est possible de mettre dans un fichier uniquement les glyphes utilisés dans le document et non pas tous les glyphes de la police, ce qui peut réduire considérablement la taille du fichier.

Puisqu'un jeu partiel est signalé en tant que tel, il est possible pour une application de fusionner des jeux partiels venant d'une même police si nécessaire.

Métadonnées

En plus des métadonnées définies par PDF 1.0, PDF 1.1 ajoute les suivantes (dans le dictionnaire des informations du document) :

- date de modification ;
- titre ;
- sujet du document ;
- mots clefs associés au document.

Articles

PDF 1.1 ajoute le concept d'articles pour être utilisé dans les cas où un document contient plusieurs flux de texte indépendants. Un article PDF identifie un de ces flux de texte, mais pas le texte lui-même, uniquement les régions sur la page qui affiche le texte de l'article. Ainsi un article est une liste ordonnée de rectangles, chacun indiquant la suite de l'article sur une page.

Présentations

Mode présentation : PDF 1.1 définit un mode de présentation de document (comme un diaporama) où il devient possible, par exemple, de défiler les pages automatiquement avec des transitions, ou de cacher des pages lors d'une présentation.

Sécurité

Cryptage : Dans la version 1.0, un document PDF était toujours lisible par n'importe qui ayant un accès au fichier. Dans la version 1.1, il devient possible de crypter les fichiers pour qu'uniquement les personnes qui ont accès à un mot de passe puissent ouvrir et lire le fichier.

Un fichier PDF crypté, n'est pas crypté entièrement. Seules les chaînes de caractères et les flux (streams) sont cryptés. La structure du fichier reste intacte.

En plus de la fonction MD5, le cryptage se fait avec l'algorithme RC4 qui est un algorithme propriété de RSA Data Security. La licence pour cet algorithme a été achetée par Adobe pour utilisation dans ses produits, mais des tiers voulant l'utiliser seront aussi obligés d'acheter la licence.

La clef était limitée à 5 octets (40 bits) pour satisfaire les restrictions d'exportation imposées par les États-Unis à l'époque. Le processus de cryptage est symétrique, le même processus étant utilisé pour décrypter.

Deux mots de passe peuvent être définis pour un document — un mot de passe propriétaire qui donne tous les droits sur le fichier, et un mot de passe utilisateur qui donne des droits limités.

Note sur la sécurité : l'un ou l'autre des deux mots de passe permet de décrypter entièrement le fichier. L'on pourrait ainsi faire tout ce que l'on veut avec seul le mot de passe utilisateur. La limitation dans les droits est donc uniquement une fonction du logiciel de lecture qui choisit de limiter les actions proposées à l'utilisateur, et non pas une limitation forcée par le format du fichier lui-même.

Annotations

PDF 1.1 ajoute un certain nombre d'améliorations par rapport à la présentation des annotations (couleur, bordure en pointillé, titre etc.).

Liens hypertexte

Liens hypertexte externes : Dans le PDF 1.0, il était possible de faire des liens hypertexte uniquement à l'intérieur d'un seul fichier PDF. Dans PDF 1.1 il devient possible de faire un lien hypertexte vers un autre fichier PDF ou vers une adresse URI.

Programmation

PDF 1.1 permet de définir des actions et qui se passent lors de l'ouverture de document ou quand on active un lien hypertexte. Ces actions sont limitées à une liste prédéfinie par PDF.

PDF v1.2

Version PDF	1.2
Version Acrobat	3.0
Date de sortie	12 novembre 1996
Identifiant PRONOM	fnt/16

Structure du fichier

Unicode : PDF 1.2 introduit la possibilité de coder des caractères en Unicode. Avant cette version le codage utilisé était un codage spécifique PDF (sur-ensemble d'ISOLatin1), mais avec le PDF 1.2 il devient possible de coder des caractères en Unicode (et donc d'utiliser les scripts comme le chinois) et aussi d'indiquer la langue utilisée.

Fichiers PDF linéarisés : Le format PDF se base sur la lecture d'une table des matières qui se trouve en fin de fichier. Tandis que cela ne pose pas de problème particulier lorsque le fichier est stocké sur un ordinateur local, ça cause bien des problèmes lorsque le fichier est stocké ailleurs sur un réseau étendu. En effet, on ne peut pas interpréter même la première page, avant d'avoir lu tout le contenu du fichier, qui peut prendre beaucoup de temps si la connexion est lente.

PDF 1.2 introduit la notion de PDF linéarisé pour ajouter des informations en début de fichier et ainsi rendre la lecture de fichier plus rapide sur un réseau étendu.

Flux externes : Il devient possible dans PDF 1.2 de créer des flux (streams) externes au fichier PDF de base. Avant, ils devaient être contenus dans le fichier lui-même.

Syntaxe : Il existe dans la version 1.2 des modifications de syntaxe dans la façon de stocker les noms dans un fichier PDF qui peut rendre des fichiers des versions antérieures illisibles. C'est notamment dû à l'utilisation du caractère '#' comme caractère d'échappement (ce n'était qu'un caractère normal avant), mais vu que *Adobe Acrobat* ne générait pas des fichiers pouvant devenir illisibles, il est peu probable que l'on rencontre des problèmes dans des fichiers PDF.

Couleur

Couleurs indexées : dans PDF 1.0 et 1.1 les couleurs indexées étaient limitées à RVB et CMJN. Avec 1.2 il est possible d'indexer tout type de couleurs sauf les motifs (patterns).

Couleurs de séparation : PDF 1.2 introduit aussi la notion d'espace de couleurs de séparation. Ce sont des couleurs définies par un nom. Si le nom est reconnu par un périphérique de sortie comme étant une encre spécifique, cette encre sera utilisée, sinon une valeur de couleur alternative est fournie pour synthétiser la couleur utilisant un des autres espaces de couleur. Aucune consigne n'est donnée sur les noms à utiliser.

Motifs : le PDF 1.2 introduit des motifs (patterns) pour créer des motifs répétitifs. Un motif peut être coloré (le motif définit sa couleur) ou non-coloré (la couleur est définie quand on applique le motif sur une page). La différence entre un motif et une forme composite est qu'une forme composite est affichée une seule fois chaque fois qu'elle est appelée, mais un motif est répété autant que nécessaire pour remplir une région.

Polices

PDF 1.2 ajoute des types de polices qui peuvent être inclus dans un fichier PDF :

- Type 0 ;
- CIDFont Type 0 (format supportant un grand nombre de glyphes, basé sur Type 1) ;
- CIDFont Type 2 (format supportant un grand nombre de glyphes, basé sur TrueType).

Impression

PDF 1.2 ajoute plusieurs fonctions qui permettent une meilleure qualité d'impression :

- ajustement de l'épaisseur des traits pour produire des traits d'épaisseur uniforme ;

Guide Méthodologique Le format PDF

- indication de surimpression lors d'impression de séparations ;
- calcul du composant noir lors de conversion RVB en CMJN ;
- calcul de la suppression de couleurs cyan, magenta ou jaune lors de conversion RVB en CMJN ;
- fonctions de transfert (pour compenser les caractéristiques des périphériques de sortie non-linéaires).

Demi-tons : le PDF 1.2 introduit la notion, déjà présente dans PostScript, de demi-tons. Un demi-ton est une façon d'imprimer un niveau de couleur par la répétition d'un motif de points. Plusieurs motifs sont définis dans la spécification.

Préférences de document

La version 1.2 introduit la notion de préférences de document pour qu'un document PDF puisse indiquer au logiciel de lecture ses préférences par rapport à la façon de disposer la fenêtre de lecture (nombre de pages visibles, autres fenêtres ouvertes, etc.).

Compression

Compression Flate : En plus des algorithmes de compression permis dans la version 1.0, la possibilité d'utiliser la compression Flate a été ajoutée dans la version 1.2. La compression Flate est similaire à LZW, mais appartient au domaine public, tandis que LZW contenait des éléments brevetés.

Formulaires interactifs « AcroForms »

À partir de PDF 1.2, PDF permet de définir des formulaires interactifs, c'est-à-dire de faire l'équivalent d'un formulaire papier où il existe des champs à remplir. (Il est à noter que dans la version anglaise on utilise le terme « form » pour indiquer à la fois les formulaires interactifs et les ressources de type forme composite. Ce sont des choses bien distinctes.).

Un seul fichier PDF peut contenir un seul formulaire, mais ce formulaire peut couvrir tout le document, et ses champs peuvent être placés sur n'importe quelle page du document.

Les types de champs suivants peuvent être définis:

- bouton (sur lequel il est possible de s'appuyer pour enclencher une action) ;
- *checkbox* (case à cocher qui peut alterner entre coché et non-coché) ;
- *radio button* (série de plusieurs options, lorsqu'on sélectionne une option les autres sont désélectionnées automatiquement) ;
- champ pour saisir du texte (une ligne ou plusieurs) ;
- liste pour le choix d'une option.

Trois actions sont permises sur un formulaire :

- soumission — le contenu du formulaire est envoyé à un URL pour traitement (soit en format HTML soit en « Forms Data Format ») ;
- réinitialisation — remettre les valeurs des champs à leur valeurs par défaut ;
- import — importer le contenu d'un fichier pour initialiser les valeurs des champs. Le fichier doit être en format « Forms Data Format ».

Le « Forms Data Format », qui est spécifique à ces fonctions, est défini dans une annexe de la spécification (son format est similaire à un fichier PDF).

Multimédia

PDF 1.2 ajoute deux types d'annotation PDF pour le multimédia :

- vidéo (movie) — pour mettre l'appel à une vidéo dans le PDF, mais le fichier vidéo est toujours externe au fichier PDF ;

- audio — pour ajouter du son dans le fichier PDF. Le fichier son peut être inclus dans le PDF ou appelé par référence à un fichier externe. Dans le cas d'un son inclus, PDF décrit exactement comment il doit être stocké — il n'est pas lié à un format spécifique. Dans le cas d'un son externe, des formats standards de l'époque (AIFF, AIFF-C, .wav, .au) sont mentionnés comme étant utilisables, mais la liste n'est pas limitative, et leur utilisation dépend des capacités du lecteur.

Programmation

Événements de curseur : Dans le PDF 1.1 on pouvait associer des actions au relâchement du bouton de la souris. Avec PDF 1.2 il devient possible de définir des actions liées aux différentes actions qui peuvent être associées aux mouvements du curseur (entrée dans une région, sortie d'une région, dépression du bouton de la souris, relâchement du bouton de la souris etc.).

OPI

OPI (*Open Prepress Interface*) est un mécanisme qui permet de mettre des images de basse résolution dans un fichier pour l'alléger lors de sa création, pour les remplacer par des images haute résolution lors de son impression finale. C'était très important aux débuts de la révolution PAO puisque les machines de l'époque s'écroulaient sous le poids de trop de fichiers à haute résolution.

PDF 1.2 définit un mécanisme dans un fichier PDF pour implémenter OPI lors de l'impression d'un fichier PDF, en ajoutant à une image les informations nécessaires à son remplacement par une version haute résolution.

PDF v1.3

Version PDF	1.3
Version Acrobat	4.0
Date de sortie	avril 1999
Identifiant	fnt/17
PRONOM	
Spécification	http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference13.pdf

Structure du fichier

Étiquettes de pages : Avant le PDF 1.3 chaque page était identifiée uniquement par sa position dans un document. Le PDF 1.3 introduit la notion d'étiquettes de pages (page labels) qui permet d'attribuer un nom à chaque page pour l'identifier.

Une étiquette de page comporte un préfixe textuel et une partie numérique. La partie numérique peut être affichée en chiffres décimaux, en chiffres romains (minuscules ou majuscules) ou comme une séquence de lettres (minuscules ou majuscules). Il est possible de définir des séquences de pages où chaque page dans la séquence utilise le même préfixe et le même type de numérotation.

Des exemples d'étiquettes pourraient être « A-1 », « A-2 », « A-3 » etc. ou « Intro-i », « Intro-ii », « Intro-iii » etc.

Dictionnaire Pièce de page (PieceInfo) : Dans PDF 1.3 un mécanisme est défini qui permet de lier des pages ou des éléments de formulaire à des données privées d'une application. (Notons par exemple que *PhotoShop* stocke ces informations de couche avec ce mécanisme).

Modélisation de l'image

Masques d'image : PDF 1.3 ajoute une fonctionnalité pour créer des masques pour des images (rastérisées).

Deux types sont possibles :

- par position — l'image de base est visible uniquement aux positions où l'image utilisée comme masque contient de la couleur ;
- par couleur — l'image de base est visible uniquement aux positions où l'image utilisée comme masque contient des couleurs dans une gamme précise, sinon ce sont les pixels de l'image masque qui sont visibles. L'effet est celui appelé incrustation (*chroma keying*) utilisé par exemple dans les bulletins de météo.

Dégradés : Jusqu'à PDF 1.2 la façon préconisée pour créer des dégradés était de créer le dégradé dans une image rastérisée et puis d'utiliser le rognage pour lui donner la forme désirée. Avec PDF 1.3 il devient possible de créer des dégradés directement et de façon beaucoup plus efficace.

Plusieurs types de dégradé peuvent être créés:

- basé sur une fonction — la fonction détermine la couleur à chaque point ;
- dégradé axial — dans un axe linéaire ;
- dégradé radial — à partir d'un point central ;
- ombrage de Gouraud en forme libre — utilise des triangles pour définir la couleur, souvent utilisé dans le rendu 3D ;
- ombrage de Gouraud en forme de réseau ;
- réseau de patch de Coons — utilise des régions entourées par quatre courbes Bézières ;
- réseau de patch de produits tensoriels — similaire aux patches de Coons sauf qu'ils utilisent 16 points de contrôle au lieu de 12.

La granularité des dégradés peut être contrôlée de façon très fine.

Images de substitution : PDF 1.3 permet d'avoir des définitions de substitution pour une image, ce qui permet dans un même document d'avoir, par exemple, une image à qualité d'écran et une deuxième à qualité impression. L'image de base est utilisée pour la visualisation à l'écran, et cette image peut avoir plusieurs images de substitution. Une seule de ces images peut être indiquée en étant l'image par défaut pour l'impression.

Fonctions : PDF 1.3 définit des nouveaux types de fonction à ceux qui existaient déjà :

- fonction à interpolation exponentielle — permet de créer des fonctions avec un calcul exponentiel et pas linéaire ;
- fonction composite à une entrée (1-input stitching function) — permet de créer une fonction à partir de plusieurs autres ;
- calcul de type PostScript — permet de créer une fonction avec un jeu limité d'instructions PostScript. PDF n'est pas un langage de programmation, mais cette fonctionnalité permet de créer des fonctions beaucoup plus riches que les simples fonctions basées sur l'interpolation.

Couleur

PDF 1.3 définit deux nouveaux espaces de couleur :

- ICCBased — définit des couleurs basées sur des profils définis par l'International Color Consortium (ICC). Pour PDF 1.3 c'est la définition ICC version 3.3 qui est retenue. Pour les cas où il n'est pas possible d'utiliser le profil ICC, il est possible dans la définition de couleur de préciser une couleur de substitution basée sur un des autres espaces de couleur (sauf un motif).
- DeviceN — l'espace de couleurs DeviceN est une généralisation de Separation qui permet pour chaque séparation de définir plusieurs colorants au lieu d'un seul. L'avantage est d'avoir plus de contrôle sur la surimpression (il est par exemple possible de contrôler les composants cyan, magenta et jaune ensemble dans une seule séparation). L'espace de couleurs Separation est effectivement un cas spécial de DeviceN où chaque séparation ne définit qu'une seule couleur.

Impression

PDF 1.3 ajoute des informations pour gérer l'impression :

- **Liste de séparations** — dans les versions précédentes de PDF on avait défini des espaces de couleur « indexé » et « séparation » qui permettent de créer des listes de couleurs applicables sur une page. Il manquait néanmoins un mécanisme pour indiquer les séparations nécessaires pour imprimer une page. PDF 1.3 ajoute un tel mécanisme pour que l'on puisse indiquer les séparations nécessaires pour l'impression d'une page utilisant l'un ou l'autre de ces deux espaces de couleur.
- **Engraissement** (trapping) — PDF 1.3 permet la création d'instructions qui définissent le grossi et le maigri. Ce sont des techniques nécessaires lorsque deux encres doivent être imprimées côte à côte pour éviter d'avoir des régions de blanc si l'impression des deux couleurs n'est pas parfaitement alignée. Il consiste à modifier la taille des régions pour créer ou non un peu de chevauchement entre les encres, mais doit bien être géré selon les différentes encres concernées pour éviter de créer des effets de bordure avec la surimpression des deux encres.

En plus des deux types de dimensions de page définis avec 1.0, le PDF 1.3 en ajoute trois autres :

- **ArtBox** — taille de la page à utiliser comme taille d'image si la page est placée comme une illustration dans un autre document ;
- **TrimBox** — la taille de la page finie (où on a l'intention de couper le papier si on avait mis d'autres indications autour). (Le MediaBox peut être plus grand puisqu'il peut comporter ces marques autour : fond perdu, ailes de coupe etc.) ;

- **BleedBox** — une région qui définit le contenu de la page, mais pas les marques d'imprimeur. Dans un environnement de production cette région est typiquement plus grande que la page par quelques millimètres et comportera les régions de fond perdu, mais pas les repères de coupe.

Demi-tons : PDF 1.3 ajoute des fonctionnalités de demi-ton par rapport à 1.2 :

- nom de demi-ton — permet de substituer une définition de demi-ton dans le PDF par une définition dans le périphérique de sortie en utilisant une correspondance de noms ;
- précision des seuils de demi-ton à 16-bits au lieu de 8.

Ordre de lecture : PDF 1.3 ajoute des informations sur l'ordre de lecture d'un document (gauche à droite ou droit à gauche) pour traiter correctement l'affichage et l'impression d'une séquence de pages.

Préférences de document

Il devient possible avec PDF 1.3 d'indiquer le niveau de zoom par défaut à appliquer à une page pour l'afficher correctement (c'est une préférence qui peut être stockée sur chaque page individuellement).

Signatures numériques

PDF 1.3 définit un type de champ dans un formulaire « AcroForm » pour inclure des signatures numériques dans un document. PDF ne définit pas le format exact que doit prendre la signature, seulement la façon de la stocker. Il est ainsi possible de stocker des signatures de type mathématique (par exemple : calcul de signature à base de clef publique et privée) ou biométrique (par exemple : signature écrite, empreinte digitale ou scan rétinale).

La spécification indique comment ajouter des types de gestionnaires de signatures. Elle mentionne plusieurs exemples de nom (Adobe.PPKLite, Entrust.PPKF, CICI.SignIt, et VeriSign.PPKVS) sans pour autant les définir ou les préconiser.

Structure logique de document

PDF 1.3 ajoute la possibilité de décrire la structure logique d'un document indépendamment de sa structure graphique.

Cette structure utilise les mêmes notions qu'utilisent XML ou SGML pour définir une structure logique et hiérarchique du document, mais la structure est stockée dans le fichier PDF comme une séquence d'objets et non pas comme un flux de type XML ou SGML. Chaque élément dans la structure a un type qui peut être lié à un type générique (titre, paragraphe, figure etc.) appelé « rôle ». Ces deux niveaux de signification (nom dans la structure, rôle) permettent de créer une structure aussi riche qu'un fichier XML basée sur une DTD spécifique, mais en même temps d'avoir une structure générique compréhensible par un lecteur qui n'a pas connaissance de la structure spécifique.

La structure logique est liée au contenu graphique de la page pour qu'il soit possible de trouver les éléments de structure à partir de l'image de la page, ou vice versa.

Annotations

PDF 1.3 ajoute plusieurs types d'annotations :

- texte libre sur la page ;
- tampon (rubber stamp) pour ajouter un effet d'ajout de tampon sur une page (comme « Approuvé », « Confidentiel », « Vendu » etc. Il est noté qu'Acrobat traite certains noms de tampon avec une présentation spéciale, même si cela ne fait pas partie de la spécification PDF ;
- trait ;
- carré ;

- cercle ;
- barré — pour barrer du texte ;
- surligné — effet de surligneur ;
- souligné — effet de soulignage ;
- encre — une série de points joints par des traits pour créer une forme prédéfinie ;
- pièce jointe — référence à un fichier embarqué dans le fichier PDF ;
- pop-up — ce type d'annotation est toujours utilisé en lien avec un autre pour permettre la modification du texte de l'autre dans une fenêtre à part.

PDF 1.3 ajoute des options d'affichage pour les annotations pour pouvoir éviter, par exemple, qu'elles soient zoomées ou mises en rotation avec la page sur laquelle elles se trouvent.

Sécurité

(Le cryptage par clef publique a été décrit dans la documentation PDF pour la première fois dans la documentation PDF 1.5, mais puisque les produits Adobe l'ont utilisé depuis PDF 1.3 il est considéré comme faisant partie de PDF 1.3.)

Cryptage : En plus du cryptage à base de mots de passe, PDF 1.3 introduit le cryptage par clefs publiques X.509 pour crypter un fichier avec une liste de destinataires. Il est possible de préciser des droits d'accès unique pour chaque destinataire. Un destinataire utilisera alors sa clef privée pour décrypter le fichier.

Programmation

PDF 1.3 permet l'utilisation de JavaScript pour définir les actions dans un document pour donner plus de possibilités que les simples options présentes dans PDF 1.2.

Fichiers embarqués

PDF 1.3 permet d'inclure un fichier de n'importe quel type à l'intérieur d'un fichier PDF et de l'extraire plus tard (même les fichiers Macintosh en deux parties). Le fichier embarqué ne participe pas au rendu du PDF, il est seulement présent en tant qu'information en plus ajoutée par l'utilisateur. Le nom de chaque fichier embarqué est également stocké dans le fichier PDF.

Il est aussi possible d'indiquer des groupes de fichiers en relation les uns avec les autres (related files) — par exemple un jeu de fichiers de séparations couleur.

Capture de pages web en PDF

PDF 1.3 introduit des structures nécessaires pour faire la capture de pages web dans un fichier PDF.

Evidemment, il était possible de créer un PDF à partir d'une page web sans modifier en rien la structure du PDF, mais il n'aurait pas été possible de retrouver les pages concernées ou de mettre à jour les contenus capturés.

Les ajouts dans la structure permettent de mémoriser l'origine des pages pour pouvoir les actualiser. Il est aussi possible de convertir des liens hypertexte vers un URL en lien hypertexte à l'intérieur du document si on inclut le document cible à l'intérieur du PDF, et puis de le reconvertir en URL si on supprime ultérieurement le document capturé.

PDF v1.4

Version PDF	1.4
Version Acrobat	5.0
Date de sortie	Acrobat : mai 2001 ; Spécification : décembre 2001
Identifiant PRONOM	fnt/18
Spécification	http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference.pdf

PDF 1.4 est la première version qui couvre toutes les possibilités de PostScript et en plus ajoute la notion de transparence qui n'existe pas dans le PostScript. On peut donc dire que le format est devenu mature et qu'à partir de cette version PDF va inexorablement remplacer PostScript dans les flux de production qui l'utilisaient.

Structure du fichier

Mise à jour incrémentale de version : Avant PDF 1.4 il était impossible à mettre à jour la version d'un fichier PDF par une mise à jour incrémentale, puisque la version était définie en début de fichier. PDF 1.4 permet aussi de mettre la version de PDF dans le catalogue du fichier (qui peut être mis à jour de façon incrémentale) et donc il devient possible de changer la version de PDF d'un fichier lors d'une mise à jour incrémentale.

Importation de fichiers PDF : PDF 1.4 définit un mécanisme par lequel il est possible d'importer le contenu d'un fichier PDF vers un autre. Avec ce mécanisme il devient possible d'inclure la définition d'une page entière d'un autre fichier PDF dans un flux et puis de le référencer à l'intérieur d'une page. Le fichier appelé peut être externe au fichier PDF ou inclus comme fichier embarqué. Il est possible de préciser une image de substitution pour les lecteurs qui ne comprennent pas ce mécanisme.

Ressources ProcSets — les ressources ProcSet n'étaient définies que pour indiquer comment générer du PostScript depuis PDF. À partir de PDF 1.4, maintenant que PDF n'est plus considéré comme simplement une façon de créer du PostScript, il est recommandé de les inclure lors de l'écriture d'un fichier PDF (pour compatibilité avec d'anciens lecteurs) mais de les ignorer à la lecture.

Modélisation de l'image

Instructions PostScript — avant PDF 1.4 la spécification de PDF n'était pas suffisamment complète pour rendre l'utilisation de PostScript inutile. La spécification de PDF 1.4 reconnaît que PDF couvre maintenant toutes les possibilités de PostScript et donc il n'y a plus besoin de recourir à des instructions PostScript. Il est donc déconseillé à partir de cette version de les inclure dans un fichier PDF avec un ou l'autre des mécanismes disponibles pour le faire.

Couleur

Intention de sortie : PDF 1.4 permet d'inclure dans un fichier PDF des indications qui permettent de faire correspondre des caractéristiques de couleur décrites dans le fichier avec celles d'un périphérique de sortie cible.

Transparence

Le langage PostScript et PDF avant 1.4 n'avaient pas de concept de transparence. Tout objet écrit sur une page cachait complètement les objets déjà placés. Avec PDF 1.4 un modèle de transparence a été introduit qui permet de créer des objets avec une transparence permettant de laisser passer plus ou moins le contenu de la page déjà présent à travers l'objet posé.

La gestion de transparence est très complexe, et un chapitre entier est dédié au sujet dans la spécification PDF 1.4.

Il est important de comprendre que tout fichier PDF antérieur ne gère pas la transparence. Par exemple, un fichier PDF/X-1 qui est basé sur le PDF 1.3 ne peut pas contenir des éléments avec une transparence — le résultat sera incorrect.

Impression

PDF 1.4 permet de préciser les paramètres de défonce ou de surimpression pour les éléments textuels individuellement.

Il introduit aussi un nouveau type d'annotation PDF pour la création de marques d'imprimeur (repères de coupe etc.).

Préférences de document

Plusieurs préférences sont ajoutées par PDF 1.4 :

- permet de mettre le titre de document dans la fenêtre qui l'affiche au lieu du nom de fichier. Il est important de souligner que c'est bien une propriété du document : la façon qu'aura un lecteur de l'interpréter dépendra bien sûr du lecteur utilisé.) ;
- permet de préciser les régions de page à utiliser pour l'affichage et l'impression. Plusieurs régions ont été définies dans les versions précédentes de PDF (MediaBox, CropBox etc.) les nouvelles préférences permettent d'explicitier certaines caractéristiques (par exemple d'avoir une région différente pour l'affichage écran que pour l'impression).

Métadonnées

Avant la version 1.4 les métadonnées d'un fichier étaient stockées dans le dictionnaire des informations du document. PDF 1.4 introduit une nouvelle façon de stocker les métadonnées dans un flux XML.

Les avantages de cette nouvelle possibilité :

- les métadonnées peuvent être attachées autant sur des composants du document que sur le document en entier. Il est donc possible de conserver, par exemple, les métadonnées d'une image incluse ;
- il devient possible pour des systèmes qui ne comprennent pas le PDF d'identifier le flux XML dans le fichier et d'utiliser les métadonnées sans avoir besoin de comprendre toute la structure du fichier.

Le flux de métadonnées est défini par une spécification appelée XMP (*Extensible Metadata Platform*) publiée par Adobe.

Sommaire

Il est possible à partir de PDF 1.4 de préciser la couleur et le style de texte pour les éléments dans la table des matières.

Présentations

(Les présentations alternatives ont été décrites dans la documentation PDF pour la première fois dans la documentation PDF 1.5, mais puisque les produits Adobe les ont utilisées depuis PDF 1.4 il est considéré comme faisant partie de PDF 1.4.)

Présentations alternatives : Le seul type de présentation alternative qui est défini est le diaporama (slideshow). Typiquement ce type de présentation est initié par l'utilisation de JavaScript sur un élément de formulaire interactif.

Cette fonctionnalité n'est pas la même chose que les présentations de type « Mode présentation » (introduites en PDF 1.1) qui traitent le document PDF dans son ensemble comme une présentation. La présentation alternative de type diaporama va présenter uniquement une série d'images stockées à l'intérieur du fichier PDF. Il n'est même pas obligatoire que ces images soient présentes sur les pages du document, mais puisque certains lecteurs PDF ne savent pas traiter une présentation alternative, il est recommandé que chaque image apparaisse dans le document pour que tous les lecteurs puissent les visualiser.

Structure logique de document

Langue globale du document : PDF 1.4 permet de préciser la langue globale du document (et aussi des éléments de texte structuré à l'intérieur du document).

PDF Balisé : PDF 1.3 avait introduit la possibilité d'inclure la structure logique du document à l'intérieur du fichier PDF, mais cette possibilité n'était pas très normalisée — chaque document pourrait décrire sa structure en utilisant des balises et des rôles différents.

PDF 1.4 évolue sur cette base en fournissant un ensemble de règles pour la description de la structure et une liste de rôles à utiliser. La description devient ainsi normalisée et donc compréhensible plus facilement par les logiciels qui lisent le PDF.

L'utilisation de PDF Balisé ne limite pas la richesse de la structure logique — il est toujours possible d'utiliser une description riche du document — mais un mécanisme est fourni pour que la richesse décrite puisse toujours être ramenée aux balises de base du PDF balisé pour permettre une compréhension par un nombre plus large de logiciels.

Les éléments de base décrivent les structures suivantes :

- paragraphes ;
- titres ;
- listes ;
- tableaux ;
- liens hypertexte ;
- illustrations.

Annotations

PDF 1.4 introduit plusieurs améliorations dans la gestion des annotations :

- chaque annotation peut avoir un nom qui l'identifie ;
- les annotations peuvent être affichées avec de la transparence ;
- un contenu alternatif d'une annotation peut être défini surtout pour l'utilisation dans les lecteurs d'accessibilité pour fournir une description textuelle de l'annotation ;
- les annotations de type texte libre peuvent préciser le cadrage à utiliser pour le texte ;
- les annotations de type trait peuvent définir des formes de fin de trait (flèche, carré etc.) ;
- les annotations de type carré et cercle peuvent préciser une couleur de remplissage.

Un nouveau type d'annotation est aussi défini : le soulignement ondulé (Squiggly underline).

Compression

Compression JBIG2 : PDF 1.4 ajoute la possibilité d'utiliser la compression JBIG2 pour les images. Cette compression ne peut être utilisée que pour les images noir et blanc (1 bit par pixel) et non pas pour les images couleur ou en niveau de gris. Ce type de compression a été défini par le *Joint Bi-Level Image Experts Group*, à l'intérieur de l'*International Organization for Standardization* (ISO). Il fournit un niveau de compression bien meilleur que la compression CCITT déjà permise dans les fichiers PDF.

Sécurité

Cryptage : Avant PDF 1.4 le cryptage était limité aux clefs de 40-bits. Avec PDF 1.4 cette limite est augmentée jusqu'à 128-bits. De plus, un nouvel algorithme est permis dans un document PDF mais il n'est pas publié pour respecter les droits d'export des États Unis.

Formulaires interactifs « AcroForms »

Plusieurs améliorations dans la gestion des formulaires interactives sont apportées par la version 1.4 :

- En plus des deux formats déjà définis (HTML ou « Forms Data Format ») il devient possible d'envoyer le contenu des champs d'un formulaire en XML ou PDF ;
- Lorsqu'on envoie le contenu des champs en « Forms Data Format » il est possible avec PDF 1.4 de préciser si on veut inclure le contenu des mises à jour incrémentales du fichier PDF ou non et si on veut envoyer le contenu des annotations ou non ;
- Des valeurs de champ pour export en Unicode peuvent être précisées.

Il existe aussi des nouveautés dans les champs possibles dans un formulaire :

- possibilité d'avoir des choix multiples dans une liste de sélection ;
- possibilité de créer un contrôle pour sélectionner un fichier.

Programmation

Il est possible de préciser des actions Javascript à effectuer lors d'opérations sur le document en entier :

- avant fermeture ;
- avant et après l'enregistrement ;
- avant et après l'impression.

Fichiers embarqués

Dans PDF 1.3 il était possible d'inclure des fichiers divers à l'intérieur du fichier PDF, mais il n'était pas facile de trouver une liste complète des fichiers embarqués.

PDF 1.4 définit un mécanisme qui permet de créer une liste des fichiers embarqués et donc de pouvoir les retrouver facilement.

PDF v1.5

Version PDF	1.5
Version Acrobat	6.0
Date de sortie	juillet 2003
Identifiant	fnt/19
PRONOM	
Spécification	http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference15_v6.pdf

Structure du fichier

Pour améliorer le niveau de compression possible dans un fichier PDF, PDF 1.5 introduit des nouveaux types de flux :

Flux d'objets (object streams) : C'est une façon de stocker les objets PDF à l'intérieur d'un objet de type flux (stream) qui permet ainsi de les compresser. C'est une façon d'améliorer la compression d'un fichier PDF tout en gardant la structure PDF intact.

On ne peut pas mettre tous les types d'objets dans un flux. Notamment on ne peut pas imbriquer des flux. On ne peut pas non plus y mettre le dictionnaire de cryptage, ni des objets qui remplacent d'autres lors d'une mise à jour incrémentale (pour peu qu'un ancien lecteur ne trouve pas la mise à jour et affiche une vieille version du document).

L'utilisation des flux d'objets est facultative. Il permet de réduire la taille du fichier, mais la lecture peut être ralentie. Il y a donc besoin de bien organiser les objets à l'intérieur de ces flux, et de regrouper les objets utilisés ensemble pour optimiser l'accès à ces objets.

Flux de références croisées (cross-reference streams) : Un deuxième mécanisme pour réduire la taille des fichiers PDF est le flux de références croisées. Le principe est le même que pour les flux d'objets — des informations qui étaient auparavant hors d'un flux, sont mises dedans pour permettre leur compression. Dans ce cas c'est la table de références croisées (cross-reference table) et le pied du fichier (footer) qui peuvent être mis dans un flux. Il devient ainsi possible de la compresser et donc réduire la taille du fichier.

Note de compatibilité : un fichier PDF qui utilise ces flux ne sera pas lisible par un lecteur antérieur à PDF 1.5 puisque le lecteur ne saura pas trouver les informations. Il est possible de créer des fichiers hybrides dans lesquels certains objets sont invisibles à des anciens lecteurs, tout en permettant la lecture du fichier, mais le fait que les objets sont cachés va forcément dégrader l'expérience utilisateur du fichier (absence de sommaire par exemple si on met les entrées de sommaire dans un flux d'objets).

Flux de texte (text streams) : PDF 1.5 ajoute la possibilité de stocker des séquences de caractères comme des flux au lieu de les stocker comme des chaînes de texte. C'est utile pour des séquences longues que l'on voudrait compresser par exemple.

Modélisation de l'image

Image : Dans PDF 1.5 une image peut-être définie avec des niveaux de couleur à 16-bits (65536 valeurs distinctes) au lieu de la limite de 8-bits (256 valeurs distinctes) dans les versions précédentes. Ça permet une meilleure restitution des niveaux de couleurs.

Signatures numériques

Il y a quelques modifications dans la gestion des signatures numériques dans le PDF 1.5. Notamment il devient possible de créer la signature sur la base de parties précisées du fichier (une liste d'objets) au lieu d'une séquence d'octets (d'habitude le fichier entier). Il devient donc possible de modifier des parties du fichier non-concernées par la signature sans invalider la signature.

Le PDF 1.5 permet aussi de stocker plus d'informations sur l'environnement lors de signature pour aider dans la validation ultérieure de la signature.

Présentations

Mode présentation : Depuis le PDF 1.1 il a été possible de définir des fonctions de transition pour utiliser le PDF à faire des présentations. PDF 1.5 améliore les possibilités. Des nouvelles transitions sont définies et il devient aussi possible de faire des transitions entre des états différents d'une page en utilisant la fonctionnalité de contenu facultatif. Il devient possible, par exemple, de cacher les éléments d'une liste sur une page et de les faire apparaître un par un.

Structure logique de document

PDF Balisé : PDF 1.5 améliore le PDF Balisé introduit dans 1.4 avec quelques évolutions pour améliorer les options d'accessibilité et corriger quelques problèmes constatés dans son utilisation :

- chaque nom d'élément dans un document doit toujours avoir un lien vers son type dans la liste des rôles, pour permettre à un élément d'avoir le même nom qu'un rôle même quand il fait la fonction d'un autre rôle ;
- ajouter des rôles pour les parties de tableau (entête, corps et pied) qui n'existaient pas dans PDF 1.4 ;
- lien entre le PDF balisé et les annotations ;
- possibilité d'ajouter des éléments Ruby et Warichu qui sont utilisés dans les textes chinois et japonais pour expliciter des mots rares ou des abréviations ;
- des attributs de présentation de tableau pour permettre à un lecteur PDF de reformater le texte sans perdre la présentation du tableau. C'est surtout utile sur les lecteurs mobiles (tablettes et téléphones) lorsque l'écran est trop petit pour lire la page PDF telle qu'elle a été formatée à l'origine.

Note : le PDF/A-1 est basé sur le PDF 1.4 et donc on ne peut pas utiliser ces améliorations dans PDF/A-1a.

Contenu facultatif

PDF 1.5 introduit la notion de contenu facultatif (optional content) qui permet de préciser que certaines parties d'un document peuvent être affichées ou cachées selon des options sélectionnées par l'utilisateur. Cette fonctionnalité est utile dans des dessins avec plusieurs couches, des cartes ou des documents multilingues par exemple.

Il est possible d'appliquer les options de contenu facultatif sur les éléments d'une page, des formes, des images et des annotations.

Bien que le contrôle des contenus facultatifs puisse être individuellement mis sur des groupes d'éléments, il existe aussi une notion d'intention pour fournir des affichages plus globaux. PDF 1.5 en définit deux : « Design » (pour représenter la vue du concepteur du document) et « View » (pour représenter la vue d'un utilisateur du document).

Note de comptabilité : puisque les lecteurs qui se basent sur le PDF avant 1.5 ne connaissent pas les instructions utilisées pour le contenu facultatif, tous les éléments du document seront toujours visibles dans un tel lecteur.

Annotations

PDF 1.5 introduit des nouveaux types d'annotations :

- polygone — une forme fermée ;
- polyligne — une ligne définie par un séries de points ;
- caret — un symbole visuel pour indiquer des modifications de texte.

Quelques nouvelles valeurs pour les terminaisons de ligne pour les annotations de type trait sont aussi ajoutées.

Dans PDF 1.5 l'ordre de navigation entre les annotations d'une page peut être défini explicitement, ce qui n'était pas le cas dans les versions précédentes.

PDF 1.5 introduit aussi la notion d'état d'annotation qui permet de les utiliser dans un workflow. Deux modèles d'états sont définis :

- Marqué — marqué ou sans marque ;
- Révision (review) — accepté, rejeté, annulé, terminé, rien.

Quelques nouvelles actions liées à des annotations peuvent aussi être définies :

- ouverture de la page qui contient l'annotation ;
- fermeture de la page qui contient l'annotation ;
- la page qui contient l'annotation devient visible ;
- la page qui contient l'annotation devient cachée.

En plus les types d'action set-state et no-op définis dans des versions antérieures de PDF sont maintenant considérés comme obsolètes.

Compression

Compression JPEG2000 : Avec PDF 1.5 il devient possible de compresser des images utilisant la compression JPEG2000. JPEG2000 est une norme internationale (ISO/CEI 15444-1) pour la compression de données d'images utilisant une transformation en ondelettes et qui donnent un meilleur taux de compression que JPEG ou CCITT. L'algorithme est utile uniquement pour la compression sans perte d'images et non pas pour d'autres types de données.

Un autre avantage de JPEG2000 est qu'il est possible de décoder l'image à une résolution plus basse (pour affichage à l'écran par exemple) sans avoir à la décoder complètement (qui serait nécessaire pour l'impression).

Portée : Le mécanisme utilisé dans un fichier PDF pour appliquer la compression est le filtre appliqué sur un flux (stream). Ceci implique que seules les informations stockées dans un flux puissent être compressées. Ce n'est pas optimal puisque ça laisse un grand nombre d'informations non-compressées dans un fichier, et donc le fichier est plus grand que nécessaire.

Pour remédier à ce problème PDF 1.5 introduit deux notions : le flux d'objets et le flux de références croisées (décrits sous « Structure du fichier »).

Sécurité

Filtres de type « Crypt » : PDF 1.5 introduit la notion de filtres de type « Crypt » qui permet de préciser les algorithmes à utiliser pour décrypter un document et les parties du document à crypter. Jusqu'à PDF 1.4 si on précisait du cryptage pour un fichier, c'était pour tout le document (les chaînes et les flux). La nouvelle fonctionnalité permet de contrôler beaucoup plus finement le cryptage : il devient possible d'utiliser des algorithmes différents pour des parties différentes du document ou d'avoir des parties cryptées et des parties non-cryptées. C'est utile par exemple si on veut crypter un fichier embarqué mais pas le fichier PDF dans son ensemble.

Permissions : PDF 1.5 inclut la possibilité de préciser des permissions sur un document sans que le document soit crypté. Auparavant il fallait crypter le document pour pouvoir préciser les droits d'accès de l'utilisateur.

Formulaires interactifs « AcroForms »

PDF 1.5 fournit quelques améliorations dans les champs d'un formulaire interactif PDF :

- possibilité de faire suivre l'état des *radio buttons* groupés entre tous les boutons du groupe ;
- possibilité de créer des champs texte de type peigne (comb) où le champ est divisé en plusieurs groupes de caractères ;
- possibilité de verrouiller des champs lorsqu'un champ de signature numérique est signé.

Texte riche : Il devient possible de mettre du texte avec mise en forme (gras, italique etc.) dans les champs d'un formulaire. Ces champs de texte riche sont basés sur l'architecture XFA (voir ci-dessous).

Note de compatibilité : les lecteurs plus anciens ne comprennent pas le texte riche, mais puisque le texte est aussi stocké sans mise en forme ils peuvent le manipuler. Le problème vient si un formulaire créé par un lecteur récent est modifié par un lecteur ancien et puis remis dans un lecteur récent — il y aura deux versions du texte (avec et sans mise en forme). Le lecteur est donc obligé de comparer le contenu textuel des deux versions pour voir lequel est le plus récent.

Formulaires interactifs XFA

PDF 1.5 introduit une deuxième façon de décrire les formulaires interactifs — les formulaires XFA. PDF 1.5 permet de décrire un formulaire complet en XFA au lieu d'utiliser les formulaires classiques « AcroForms » qui existent depuis PDF 1.2.

Quand on crée un formulaire en PDF à partir de PDF 1.5, il existe donc deux façons de le faire. Les formulaires de type « AcroForms » sont mieux intégrés dans PDF mais moins souples, les formulaires XFA sont plus dynamiques mais mal intégrés dans le PDF dans son ensemble.

Multimédia

PDF 1.5 augmente beaucoup les possibilités d'inclusions d'éléments multimédia dans un fichier PDF. Notamment il devient possible :

- d'embarquer des fichiers multimédias de types divers dans le fichier PDF. (Depuis PDF 1.2 l'utilisation de vidéo était limitée à des appels à des fichiers externes) ;
- de placer des lecteurs de médias divers sur la page ou dans une fenêtre à part pour visualiser les contenus multimédias ;
- de préciser plusieurs formats d'un même contenu multimédia pour permettre un choix de visualisation lors de la lecture du contenu (pour prendre en compte la bande passante disponible par exemple) ;
- de préciser les préférences pour le format à utiliser lorsque plusieurs sont fournis et pour le logiciel à utiliser pour la lecture.

Note : les annotations PDF de type vidéo (movie) sont maintenant considérées comme obsolètes, et sont remplacées par les annotations PDF de type écran (screen).

PDF v1.6

Version PDF	1.6
Version Acrobat	7.0
Date de sortie	janvier 2005
Identifiant PRONOM	fnt/20
Spécification	http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference16.pdf

Modélisation de l'image

Taille de page PDF : La taille d'une page PDF est limitée par le nombre maximum qui peut être exprimé dans PDF pour décrire le nombre d'unités (en 1/72 pouces) que comporte une dimension de page. PDF 1.6 permet de redéfinir les unités de base en multiples de 1/72 pouces pour pouvoir agrandir la taille de page. La résolution possible de positionnement des éléments sur la page sera réduite par la même proportion.

Couleur

La définition de l'espace de couleurs DeviceN a été améliorée en PDF 1.6 pour créer ce qu'on appelle des espaces de couleur NChannel. Ils donnent plus de souplesse aux applications dans la représentation des couleurs qui ne sont pas disponibles sur le périphérique cible, en les laissant choisir la façon de mélanger les couleurs au lieu d'être restreints par les fonctions de teinte fournies lorsqu'on ne peut pas imprimer directement avec les composants (par exemple pour l'affichage à l'écran).

Polices

Avant PDF 1.6, pour embarquer une police OpenType dans un fichier PDF il était nécessaire de le convertir en TrueType ou Type 1 (vu que OpenType décrit ses glyphes dans ces deux formats ce n'était pas excessivement difficile). Avec PDF 1.6 il devient possible d'embarquer une police OpenType directement. OpenType est une extension de TrueType qui permet l'utilisation du Compact Font Format (CFF). Cette amélioration permet aussi d'inclure les données de substitution de glyphes, d'approches et d'ajustement des lignes de base.

Impression

Échelle : PDF 1.6 ajoute la possibilité de préciser l'échelle à utiliser par défaut lors de l'impression d'un document.

Filigranes : PDF 1.6 introduit un type d'annotation PDF appelé filigrane (watermark) pour positionner des éléments sur une page de telle façon qu'ils aient toujours la même position et mêmes dimensions indépendamment de la taille de la page imprimée.

Signatures numériques

PDF 1.6 ajoute des améliorations et des clarifications dans l'utilisation de signatures numériques :

- changement dans la façon de valider une signature MDP (détection et prévention des modifications), aussi connue comme signature auteur ou signature de certification ;
- stockage des détails de l'horodatage de la signature et l'adresse du serveur qui a fait l'horodatage.

Pour les signatures qui utilisent des clefs publiques/privées RSA, PDF 1.6 ajoute la possibilité d'utiliser l'algorithme SHA256 (256-bits) pour calculer l'empreinte (digest) pour les signatures PKCS#7. (Cette

fonctionnalité n'est pas mentionnée dans la spécification du PDF 1.6, mais la spécification PDF 1.7 indique qu'elle a été ajoutée dans la version 1.6.)

Structure logique de document

PDF Balisé : PDF 1.6 introduit une façon d'indiquer dans le texte balisé si l'ordre des éléments n'est pas certain. C'est utile lorsqu'on crée le PDF depuis une application qui ne fournit pas assez d'informations pour être sûr de l'ordre logique des éléments.

Il devient aussi possible dans les éléments de groupe d'indiquer si leur contenu est divisé en colonnes et d'en donner leur nombre et leurs dimensions.

Contenu facultatif

Dans le PDF 1.6 deux améliorations sont apportées par rapport au contenu facultatif introduit en PDF 1.5 :

- possibilité de préciser une expression booléenne arbitraire pour déterminer la visibilité d'un contenu ;
- possibilité de verrouiller le changement d'état d'un contenu facultatif par l'interface utilisateur (il reste quand même possible de changer l'état par Javascript).

Sécurité

Cryptage : Les possibilités de cryptage sont améliorées en PDF 1.6 pour permettre l'utilisation de l'algorithme AES (Advanced Encryption Standard).

Illustrations 3D

PDF 1.6 permet d'inclure des images en 3D décrites avec le format U3D. Il devient possible :

- de mettre des objets 3D sur la page (et pas seulement dans une fenêtre à part ou dans une autre interface) ;
- plusieurs vues d'un objet peuvent être incluses ;
- des éléments 2D peuvent être superposés sur des objets 3D ;
- les pages qui contiennent des objets 3D peuvent être imprimées ;
- il est possible de changer l'angle et l'orientation de visualisation des objets 3D ;
- des scripts JavaScript peuvent manipuler les objets 3D.

La gestion des objets 3D se fait avec :

- les annotations PDF 3D qui fournissent une caméra virtuelle ;
- les flux 3D qui décrivent les objets 3D. Dans PDF 1.6 cette description se fait avec le format U3D ;
- les vues 3D qui précisent la relation entre la caméra et les objets 3D ;
- les systèmes de coordonnées 3D.

Annotations

PDF 1.6 introduit plusieurs améliorations au niveau des annotations :

- possibilité de grouper les annotations pour qu'elles fonctionnent ensemble lorsque l'utilisateur les manipule ;
- pour les annotations de type texte libre :
 - possibilité de définir un trait (avec coude si voulu) entre une annotation et l'endroit dans le texte auquel elle se réfère ;
 - des sous-types définis par les intentions (indications du comportement de l'annotation) FreeTextCallout et FreeTextTypeWriter.
- pour les annotations de type trait :

- possibilité d'ajouter des traits orthogonaux d'une longueur définie aux extrémités d'une annotation de type trait ;
- possibilité d'ajouter un libellé sur les annotations de type trait ;
- des sous-types définis par les intentions (indications du comportement de l'annotation) LineArrow (flèche) ou LineDimension (indication de dimension) ;
- une terminaison de trait supplémentaire (trait oblique).
- pour les annotations de type polygone :
 - un sous-type PolygonCloud (objet nuage).

Liens hypertexte

Avec PDF 1.6, il devient possible de définir une région non-rectangulaire (mais définie par quatre points) comme région active d'un lien hypertexte ; et aussi de faire un lien hypertexte entre le PDF principal et des fichiers PDF inclus comme fichiers embarqués.

CAO

Propriétés de mesure : PDF 1.6 ajoute la possibilité d'indiquer les vraies dimensions d'éléments contenus sur une page PDF. Cette fonctionnalité est surtout utile dans les métiers qui utilisent la CAO où des images dessinées sur la page correspondent à des objets dans le monde réel. Il devient ainsi possible de fournir des indications des vraies dimensions des objets décrits au lieu d'avoir accès uniquement aux dimensions des traits sur la page.

Propriétés utilisateur : La plupart des propriétés des éléments dans une page décrivent la présentation de l'élément (couleur, style etc.). Les propriétés utilisateur, introduit avec PDF 1.6, permet de stocker d'autres informations sur un élément. Cette fonctionnalité vise surtout la CAO où des objets peuvent avoir la même apparence physique, mais des caractéristiques différentes qui ne correspondent pas à leur apparence (comme type ou numéro de référence par exemple).

PDF v1.7 (ISO 32000-1)

Version PDF	1.7 (ISO 32000-1)
Version Acrobat	8.0
Date de sortie	novembre 2006
Identifiant PRONOM	fnt/276
Spécification	http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_1-7.pdf

PDF 1.7 est la première version de PDF à être normalisée. La spécification de PDF 1.7 publiée par Adobe correspond exactement à la norme ISO 32000-1.

Structure du fichier

Besoins du document : PDF 1.7 introduit un dictionnaire des besoins d'un document pour indiquer à un lecteur ce qui est nécessaire à une lecture complète du document ; ceci permet au lecteur d'indiquer s'il est capable ou non de tout reproduire.

La norme définit une structure générale pour indiquer les besoins du document avec une série de critères, mais un seul critère est défini dans PDF 1.7 : la nécessité ou non de pouvoir exécuter des scripts JavaScript.

La méthode JavaScript qui assure cette même fonctionnalité (le gestionnaire des besoins) est décrite en indiquant que c'est pour garder la compatibilité avec des fichiers existants. Puisque il n'y avait pas de mention dans les spécifications des versions précédentes, on peut supposer que c'était une fonctionnalité Adobe qui n'était pas documentée.

Extensions : PDF 1.7 définit un mécanisme pour créer des extensions à PDF tout en restant compatible avec le PDF de base. Ce mécanisme était devenu nécessaire pour permettre à Adobe de continuer à ajouter des fonctionnalités à PDF tout en laissant à ISO le contrôle du format.

Préférences de document

Dans PDF 1.7 plusieurs options d'impression peuvent être stockées dans les préférences de document :

- si le document doit être imprimé recto/verso ou non, et si oui, l'orientation de la page ;
- si la taille de page doit être utilisée pour sélectionner le bac de papier ;
- la liste des pages du document à imprimer par défaut ;
- le nombre de copies du document à imprimer par défaut.

Signatures numériques

Pour les signatures qui utilisent des clefs publiques/privées RSA, PDF 1.7 ajoute plusieurs possibilités d'algorithme pour calculer l'empreinte (digest) pour les signatures PKCS#7 :

- SHA384 (384-bits);
- SHA512 (512-bits);
- RIPEMD160 (128-bits).

PDF 1.7 ajoute aussi plusieurs façons d'ajouter des exigences par rapport au contenu de la signature et à sa validation.

Annotations

Il devient possible avec PDF 1.7 de verrouiller une annotation pour empêcher l'utilisateur de modifier son contenu (ce qui ne l'empêche pas de la supprimer).

Il existe des nouvelles possibilités avec PDF 1.7 :

- Les annotations de type trait :
 - un nom qui indique la position (inline ou top) ;
 - un dictionnaire de mesures pour préciser l'échelle et les unités utilisées par le trait peut y être attaché;
 - positionnement du libellé si précisé.
- les annotations de type ligne et polyligne :
 - la notion d'intention pour les annotations est élargie aux annotations de type polyligne (pour indiquer qu'il s'agit d'une dimension) et les polygones peuvent aussi être indiqués comme étant des indications de dimension ;
 - un dictionnaire de mesures pour préciser l'échelle et les unités utilisées par le trait peut y être attaché.
- PDF 1.7 introduit les annotations de type biffure (en anglais « *redaction annotations* »). Il ne faut pas confondre le terme anglais « *redaction* » avec le mot français « rédaction ». En anglais le mot indique le processus éditorial qui consiste à combiner plusieurs manuscrits en un seul. Les annotations de type biffure ont pour but d'indiquer des morceaux de contenu à supprimer. Ainsi un utilisateur applique des annotations de type biffure aux lieux où il veut supprimer du contenu, et suite à cela, l'application supprime le contenu marqué et le remplace par un marqueur qui indique la suppression de contenu.

Formulaires interactifs XFA

Dans PDF 1.7 il est possible d'indiquer s'il faut régénérer l'affichage d'un formulaire XFA à l'ouverture d'un document.

Illustrations 3D

PDF 1.7 ajoute des possibilités dans l'utilisation 3D :

- Il est possible de préciser comment des illustrations 3D doivent être dessinées, coloriées et illuminées, et comment faire des coupes transversales, le tout sans l'utilisation de JavaScript ;
- Il devient possible de préciser des informations d'état sur des nœuds comme visibilité, opacité, position et orientation ;
- Il permet aussi la lecture d'animations 3D avec la précision de style et d'échelle de temps sans script ;
- L'activation d'une illustration 3D peut déclencher l'affichage d'éléments spécifiques dans l'interface utilisateur (comme une barre d'outils) ;
- Il devient possible d'appliquer des annotations 2D sur des vues spécifiques 3D. Auparavant les annotations étaient positionnées à un endroit spécifique sur une page. Il devient donc possible de les positionner sur une vue spécifique 3D.

Fichiers embarqués

Avant PDF 1.7, les noms de fichiers dans une spécification de fichier étaient de simples chaînes de caractères 8-bit. PDF 1.7 introduit la possibilité de stocker ces noms en Unicode et permettre ainsi des noms multilingues.

La spécification séparée de noms de fichier DOS, Mac et Unix est maintenant considérée comme étant obsolète.

Collection de fichiers embarqués : il devient possible de créer des collections de fichiers embarqués ce qui permet d'indiquer au lecteur comment présenter la liste des fichiers embarqués dans l'interface utilisateur.

PDF v1.7 niveau d'extension 1

Version PDF	1.7 niveau d'extension 1
Version Acrobat	8.1
Date de sortie	avril 2007
Identifiant PRONOM	—

Puisque avec la version 1.7 de PDF, Adobe a lâché le contrôle du format en faveur de l'ISO, dès qu'il y avait la nécessité d'ajouter des fonctionnalités PDF pour une nouvelle version d'Acrobat, il n'était plus possible de créer une nouvelle version de PDF. Adobe a donc choisi d'utiliser le mécanisme d'extensions.

Les modifications décrites ici ne sont donc pas automatiquement reportées dans la version ultérieure de PDF. Adobe les propose à l'ISO, mais leur inclusion ou non n'est plus sous le contrôle d'Adobe.

Formulaires interactifs XFA

Cette extension améliore les possibilités de texte riche dans les formulaires interactives XFA en utilisant les versions de XFA 2.5 (niveau d'extension 1) et 2.6 (niveau d'extension 2).

Illustrations 3D

Avec cette extension il devient possible d'utiliser PRC comme description 3D au lieu de U3D qui existait dans PDF depuis le PDF 1.6.

PRC est un format de description 3D qui est hautement compressé.

Cette extension permet aussi l'utilisation de U3D (ECMA-363) troisième édition au lieu de la première édition permise par PDF 1.6.

PDF v1.7 niveau d'extension 3

Version PDF	1.7 niveau d'extension 3
Version Acrobat	9.0
Date de sortie	juin 2008
Identifiant PRONOM	—

(Le niveau d'extension 2 n'est pas publié et ne correspond pas à un produit vendu.)

Préférences de document

Cette extension permet de forcer la valeur de certaines des préférences de document de manière à ce qu'elles ne puissent pas être modifiées. C'est utile, par exemple, pour les documents d'ingénierie et d'architecture pour empêcher un document d'être imprimé à une échelle autre que celle prévue par l'auteur.

Sigatures numériques

Contraintes : Cette extension ajoute la possibilité de préciser des informations de contrainte lors de l'application d'une signature numérique

- LockDocument pour limiter l'interface utilisateur après application de la signature ;
- AppearanceFilter pour filtrer l'apparence des signatures lors de la signature et la certification du document par nom.

Structure logique de document

En plus des identifiants d'artefact déjà définies, il devient possible avec l'extension 3 de définir un artefact de type BatesN pour identifier un numéro Bates.

Les numéros Bates sont simplement une façon d'identifier chaque page d'une collection de documents individuellement comme on faisait autrefois avec une machine à tamponner un numéro unique sur chaque page.

Annotations

Cette extension ajoute deux possibilités pour préciser dans un document PDF l'ordre d'activation des annotations lorsqu'un utilisateur utilise le clavier pour se déplacer entre les annotations :

- ordre défini dans le fichier (mélange d'annotations et de champs des formulaires interactifs) ;
- champs de formulaire suivis par d'autres annotations.

Sécurité

Cryptage : Cette extension ajoute un nouvel algorithme pour la création de mots de passe avec l'utilisation d'Unicode.

Il y a trois parties concernées :

- une nouvelle variante de AES (AESV3) qui utilise le cryptage AES-256 ;
- un nouveau filtre de cryptage qui permet des clefs de 256-bits (32 octets) et ne mélange pas l'identifiant objet avec la clef de cryptage ;
- une nouvelle valeur de révision de cryptage (5) active un nouvel algorithme de cryptage.

Formulaires interactifs « AcroForms »

Codes-barres : Cette extension permet un nouveau type de champ dans les formulaires interactifs pour les codes-barres. Ces champs permettent d'afficher un code-barres généré depuis le contenu textuel d'un autre champ.

Trois types de code-barres sont permis :

- PDF417 (ISO/IEC 15438) — un code-barres à deux dimensions. Le PDF signifie *Portable Data File* et n'a rien à voir avec le format PDF décrit dans ce document ;
- QR Code (ISO/IEC 18004) — un code-barres à deux dimensions (Quick Response code) ;
- Datamatrix (ISO/IEC 16022) — un code-barres à deux dimensions.

La taille du code-barres peut être précisée.

Formulaires interactifs XFA

Cette extension permet d'enregistrer les données d'un formulaire XFA dans le dictionnaire des noms pour le rendre compatible à PDF/A-2.

Multimédia

Cette extension ajoute un type d'annotation PDF qui s'appelle « RichMedia ». Il permet d'inclure dans un document PDF des applications Flash, des vidéos, des contenus audio et d'autres contenus multimédia. Il incorpore la structure d'annotation 3D existante pour supporter la vidéo Flash et des variantes compatibles H.264.

Cette annotation permet un lien entre les scripts Flash et les scripts Acrobat.

Il devient possible de lier des actions à des positions de chapitre dans les vidéos.

Illustrations 3D

Mesures : Cette extension permet de stocker dans un fichier PDF des annotations et des indications de distance sur les illustrations 3D.

Fichiers embarqués

Collection de fichiers embarqués : Cette extension ajoute plusieurs possibilités dans la gestion des collections de fichiers embarqués :

- l'utilisation de SWF pour créer des « navigateurs » — des interfaces customisées pour fournir l'accès aux éléments de la collection ;
- possibilité de diviser horizontalement ou verticalement la vue de la collection pour afficher soit les détails des fichiers, soit des petits icônes ;
- possibilité d'indiquer des vignettes pour chaque fichier d'une collection ;
- possibilité d'indiquer la taille compressée des fichiers de la collection ;
- possibilité d'indiquer les couleurs qu'un navigateur doit utiliser pour afficher la collection ;
- possibilité d'organiser les fichiers dans une structure hiérarchique de dossiers.

Informations géomatiques

Cette extension ajoute la possibilité de lier les informations dans un fichier PDF à de vraies coordonnées sur la terre. Il permet par exemple d'identifier la position géographique d'une maison depuis les plans enregistrés en PDF.

Guide Méthodologique

Le format PDF

Le système de coordonnées projetées utilisé peut être décrit avec un code de référence EPSG (<http://www.epsg.org/>) ou par un identifiant comme décrit par le document 01-009 du *Open Geospatial Consortium*.

PDF v1.7 niveau d'extension 5

Version PDF	1.7 niveau d'extension 5
Version Acrobat	9.1
Date de sortie	2009
Identifiant PRONOM	—

Cette extension est très limitée. Pour la plupart il corrige quelques erreurs dans les descriptions de fonctionnalités déjà présentes ou clarifie leur comportement.

Formulaires interactifs « AcroForms »

L'utilisation de liens hypertexte est permise dans les champs de type texte riche.

Formulaires interactifs XFA

Cette extension améliore les possibilités de texte riche dans les formulaires interactives XFA en utilisant la version 3.0 de XFA.

PDF v2.0 (ISO 32000-2)

Version PDF	2.0 (ISO 32000-2)
Date de sortie	(future)
Identifiant PRONOM	—

La norme ISO 32000-2 n'est pas encore terminée, et la liste définitive des modifications n'est pas arrêtée. On fournit ici simplement une liste des changements proposés en décembre 2011 à titre indicatif.

Structure du fichier

Besoins du document : Le mécanisme introduit dans PDF 1.7 pour indiquer les fonctionnalités nécessaires pour lire le document complètement (qui était limité au besoin de scripting) a été augmenté dans PDF 2.0 pour indiquer bien plus complètement les besoins avec toute une liste des fonctionnalités qui peuvent être nécessaires pour un document PDF.

Besoins du lecteur : PDF 2.0 introduit une liste des exigences qu'un fichier PDF peut imposer à un lecteur dans la façon de traiter son contenu. Il ne faut pas confondre les besoins du document (liste de fonctionnalités — le support de XFA par exemple) avec les besoins du lecteur (liste d'exigences dans le traitement — le respect strict des séparations de couleur par exemple).

Parties de document : PDF 2.0 introduit la notion de parties de document (DPart). C'est une façon d'identifier une page ou une séquence de pages comme étant une unité de traitement, et a été introduit surtout pour les besoins de PDF/VT (l'éditique).

Couleur

Intention de sortie : PDF 2.0 augmente les possibilités introduites dans PDF 1.4 sur les intentions de sortie en permettant de les inclure sur des pages individuelles et pas juste comme paramètres de fichier global comme c'était le cas auparavant. Un nombre d'extensions dans la description des intentions ont aussi été ajoutés.

Compensation du point noir : PDF 2.0 ajoute la possibilité d'indiquer sur les éléments de la page s'il faut ou non utiliser la compensation du point noir lorsqu'on fait des conversions à base de couleur CIE.

Préférences de document

PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui permet de forcer la valeur de certaines des préférences de document de telle manière à ce qu'elles ne puissent pas être modifiées. C'est utile, par exemple, pour les documents d'ingénierie et d'architecture pour empêcher un document d'être imprimé à une échelle autre que celle prévue par l'auteur.

Signatures numériques

Types : En PDF 2.0 les signatures à base du *Cryptographic Message Syntax* (CMS) et de CADES (CMS *Advanced Electronic Signatures*) sont permises.

Lieu de stockage des éléments de sécurité du document : PDF 2.0 introduit la notion de lieu de stockage des éléments de sécurité du document.

Il a été reconnu qu'après une période de temps (comme 5 ans par exemple) il pourrait être difficile de valider une signature numérique parce que les éléments utilisés pour la valider (comme des horodatages ou des listes de révocation) auraient passé leur date limite de validité.

Pour essayer de faciliter la validation d'une signature à long terme, il devient donc possible de stocker ces éléments dans le lieu de stockage des éléments de sécurité du document, à l'intérieur du fichier PDF, pour qu'ils soient disponibles pour valider la signature exactement comme c'était fait lorsque la signature a été créée, ou peu de temps après.

Horodatage au niveau document : PDF 2.0 ajoute la possibilité de créer des horodatages au niveau document qui est une méthode de stocker quel était le contenu exact du fichier à un moment donné en prenant l'empreinte du fichier, comme pour la signature numérique, et la fournir à l'autorité d'horodatage pour créer une preuve du contenu au moment précis.

Contraintes : PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui ajoute la possibilité de préciser des informations de contrainte lors de l'application d'une signature numérique :

- LockDocument pour limiter l'interface utilisateur après application de la signature ;
- AppearanceFilter pour filtrer l'apparence des signatures lors de la signature et la certification du document par nom.

Structure logique de document

PDF Balisé : PDF 2.0 introduit un nombre de nouvelles balises à utiliser en standard dans un document. À l'heure de l'écriture de ce rapport, ce travail n'est pas terminé et il existe un nombre de questions à résoudre sur la façon dont le PDF Balisé va évoluer pour prendre en compte les éléments complexes comme les formules mathématiques ou chimiques.

Annotations

Ordre d'activation : PDF 2.0 reprend du niveau d'extension 3 les deux possibilités supplémentaires pour préciser dans un document PDF l'ordre d'activation des annotations lorsqu'un utilisateur utilise le clavier pour se déplacer entre les annotations :

- ordre défini dans le fichier (mélange d'annotations et de champs des formulaires interactifs) ;
- champs de formulaire suivi par d'autres annotations.

Transparence : il devient possible avec PDF 2.0 de définir une transparence pour les annotations.

Améliorations particulières des annotations :

- pour les annotations de type tampon ;
- des sous-types définis par les intentions pour indiquer si le tampon vient d'un contenu PDF, si c'est une image, ou si ça reproduit un tampon à l'encre.

Liens hypertexte

PDF 2.0 crée un nouveau type de lien hypertexte qui a comme cible une partie de document (DPart).

Sécurité

Cryptage : PDF 2.0 permet l'utilisation de filtres de type « Crypt » qui utilisent une clef de longueur de 256-bits. Il est similaire à la fonctionnalité décrite dans le niveau d'extension 3 d'Adobe, mais pas identique. La valeur de hash utilisé dans le niveau d'extension 3 était simplement une valeur SHA-256. Dans PDF 2.0 la valeur de hash est calculée par un algorithme fourni dans la spécification. Les deux méthodes sont donc incompatibles.

La méthode Adobe de l'extension niveau 3 est signalée comme étant obsolète, et n'aura jamais fait partie d'une norme ISO. Le numéro de version du gestionnaire de sécurité utilisé par Adobe dans le niveau 3 n'est pas inclus dans la séquence des numéros valables définie par la norme ISO ; si un fichier utilisant la gestionnaire Adobe est trouvé il sera donc possible de l'identifier par ce numéro de version.

Formulaires interactifs « AcroForms »

Codes-barres : PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui permet un nouveau type de champ dans les formulaires interactifs pour les codes-barres. Ces champs permettent d'afficher un code-barres généré depuis le contenu textuel d'un autre champ.

Trois types de code-barres sont permis :

- PDF417 (ISO/IEC 15438) — un code-barres à deux dimensions. Le PDF signifie *Portable Data File* et n'a rien à voir avec le format PDF décrit dans ce document ;
- QR Code (ISO/IEC 18004) — un code-barres à deux dimensions (Quick Response code) ;
- Datamatrix (ISO/IEC 16022) — un code-barres à deux dimensions.

La taille du code-barres peut être précisée.

Formulaires interactifs XFA

PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui permet d'enregistrer les données d'un formulaire XFA dans le dictionnaire des noms pour le rendre compatible à PDF/A-2.

Il ajoute aussi, de façon similaire, un mécanisme pour stocker des images et des ressources Flash par nom pour utilisation par les formulaires interactifs XFA.

Multimédia

PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui ajoute un type d'annotation PDF qui s'appelle « RichMedia ». Il permet d'inclure dans un document PDF des applications Flash, des vidéos, des contenus audio et d'autres contenus multimédia. Il incorpore la structure d'annotation 3D existante pour supporter la vidéo Flash et des variantes compatibles H.264.

Illustrations 3D

PRC : PDF 2.0 reprend la fonctionnalité du niveau d'extension 1 d'Adobe qui permet d'utiliser PRC comme description 3D au lieu de U3D qui existait dans PDF depuis le PDF 1.6.

U3D : PDF 2.0, comme le niveau d'extension 1 d'Adobe permet l'utilisation de U3D (ECMA-363) troisième édition au lieu de la première édition permise par PDF 1.6.

Mesures : PDF 2.0 reprend la fonctionnalité du niveau d'extension 3 d'Adobe qui permet de stocker dans un fichier PDF des annotations et des indications de distance sur les illustrations 3D.

Programmation

Dans PDF 2.0, la façon définie précédemment pour lancer des applications externes qui fournissait des actions spécifiques pour les plateformes Windows, Macintosh et Unix a été identifiée comme obsolète et ne devrait plus être utilisée.

Fichiers embarqués

Collection de fichiers embarqués : PDF 2.0 reprend du niveau d'extension 3 d'Adobe plusieurs possibilités dans la gestion des collections de fichiers embarqués :

- l'utilisation de SWF pour créer des « navigateurs » — des interfaces customisées pour fournir l'accès aux éléments de la collection ;

Guide Méthodologique

Le format PDF

- possibilité de diviser horizontalement ou verticalement la vue de la collection pour afficher soit les détails des fichiers, soit des petits icônes ;
- possibilité d'indiquer des vignettes pour chaque fichier d'une collection ;
- possibilité d'indiquer la taille compressée des fichiers de la collection ;
- possibilité d'indiquer les couleurs qu'un navigateur doit utiliser pour afficher la collection ;
- possibilité d'organiser les fichiers dans une structure hiérarchique de dossiers.

Fichiers associés : Avant PDF 2.0 on pourrait inclure des fichiers dans le fichier PDF comme des fichiers embarqués, mais aucune indication n'était donnée par rapport à leur lien avec le fichier PDF ni pourquoi ils étaient inclus.

Dans PDF 2.0 un mécanisme est défini pour lier le contenu du fichier PDF avec les fichiers embarqués (pour indiquer un fichier source du document, un fichier alternatif pour un tableau etc.).

Ce mécanisme est identique à celui introduit dans PDF/A-3 (qui est basé sur PDF 1.7) et sert à le rendre plus générique que le PDF/A-3 seul.

Informations géomatiques

PDF 2.0 reprend du niveau d'extension 3 d'Adobe la possibilité de lier les informations dans un fichier PDF à de vraies coordonnées sur la terre. Il permet par exemple d'identifier la position géographique d'une maison depuis les plans enregistrés en PDF.

Le système de coordonnées projetées utilisé peut être décrit avec un code de référence EPSG (<http://www.epsg.org/>) ou par un identifiant comme décrit par le document 01-009 du Open Geospatial Consortium.

Les Normes PDF

PDF/A

Les normes PDF/A ont pour but de proposer un sous-ensemble des fonctions PDF que l'on peut considérer comme étant toujours disponibles dans les lecteurs futurs et donc de définir un format de fichier pérenne.

Chaque norme PDF/A est conçue indépendamment des autres pour définir un format distinct. Ces normes ne seront pas mises à jour ni suspendues. L'idée pour l'archivage est que ces normes restent valables sans modification (sauf d'éventuelles corrections d'erreurs dans le texte ou la clarification d'ambiguïtés).

L'idée est que si une norme PDF/A contient toutes les fonctionnalités nécessaires pour un document, il n'y aura jamais besoin de migrer le document vers une nouvelle norme parce que toutes les normes PDF/A seront toujours pérennes. Par contre si un document a besoin de la transparence (par exemple), PDF/A-1 n'est pas suffisant et PDF/A-2 sera nécessaire.

Le choix de la norme à utiliser est donc fait en fonction des normes disponibles au moment de sa création et des fonctionnalités utilisées dans le document. Un document utilisant la transparence ne pouvait pas être enregistré avec PDF/A-1 et donc n'était pas archivable selon ces normes avant la sortie de PDF/A-2.

PDF/A-1

Numéro de norme	ISO 19005-1
Date de publication	1 ^{er} octobre 2005
Version corrigée	1 ^{er} décembre 2005
Titre	Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme — Partie 1: Utilisation du PDF 1.4 (PDF/A-1)
Version de base de PDF	1.4
Identifiant PRONOM	PDF/A-1a : fmt/95 PDF/A-1b : fmt/354

Les buts proclamés de cette norme :

- Définir un format de fichier basé sur le PDF pour la représentation de documents dans une manière à préserver leur apparence visuelle avec le temps, indépendamment des outils utilisés pour les créer, les stocker ou les visualiser ;
- Fournir un cadre pour l'enregistrement du contexte et de l'historique de ces documents avec l'utilisation de métadonnées à l'intérieur des fichiers ;
- Définir un cadre pour la représentation de la structure logique et d'autres informations sémantiques à l'intérieur de ces fichiers.

La norme décrit seulement le format des fichiers et ne couvre pas les autres aspects d'archivage tel que le stockage des fichiers ou la validation que le fichier créé est bien une copie conforme du document originel.

Niveau de conformité

La norme définit deux niveaux de conformité qu'un fichier peut avoir par rapport à la norme :

- Niveau A (avancé) — le fichier est conforme à toutes les exigences de la norme. Un fichier avec ce niveau de conformité est identifié comme étant du PDF/A-1a ;
- Niveau B (basic) — le fichier est conforme à toutes les exigences de la norme par rapport à l'aspect visuel du document, mais ne contient pas les informations de contenu textuel ou de structure logique du document. Un fichier avec ce niveau de conformité est identifié comme étant du PDF/A-1b.

Structure du fichier

Linéarisation : La linéarisation d'un fichier PDF/A-1 est permis, mais pas obligatoire. Un lecteur conforme PDF/A-1 est censé ignorer les informations de linéarisation.

Dépendances externe : Il est possible avec PDF de définir des dépendances du fichier PDF avec des fichiers externes. Ces dépendances sont interdites par PDF/1-A pour que le fichier PDF/A-1 soit complètement autonome.

Limites dans l'implémentation : La spécification de PDF indique un certain nombre de limites dans les tailles et les quantités des éléments qui sont contenus dans un fichier PDF. Dans cette spécification ces limites sont données à titre indicatif par rapport aux valeurs utilisées par Acrobat, mais le respect de ces limites n'est pas exigé. PDF/A-1 rend le respect de ces limites obligatoire.

Modélisation de l'image

Images : Chaque image dans un fichier PDF/A-1 doit avoir une seule représentation (il est possible avec PDF d'en spécifier plusieurs) et il n'est pas permis de laisser au lecteur le soin d'interpoler les images pour améliorer leur rendu (puisque cela aurait l'effet de rendre l'image de façon différente selon l'algorithme utilisé par chaque lecteur).

PostScript : Il est interdit d'inclure dans un fichier PDF/A-1 des instructions PostScript. C'était permis dans les premières versions de PDF, mais fortement déconseillé depuis le 1.4.

Couleur

Espaces de couleur : Dans un fichier PDF/A-1, toutes les couleurs doivent être définies de façon indépendante de la sortie. Si des couleurs dépendantes de la sortie sont utilisées, une intention de sortie doit être fournie pour permettre leur interprétation correcte, et on ne peut pas utiliser des couleurs RVB et CMJN dans le même document.

Pour toute couleur qui se base sur un profil ICC, le profil associé doit être embarqué dans le fichier PDF.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/A-1.

Polices

Les polices utilisées dans un document PDF/A-1 doivent être conforme à PDF 1.4. Toute police utilisée dans le document doit être embarquée dans le fichier PDF/A-1. Il est interdit d'utiliser des polices dont la licence ne permet pas de les embarquer de façon à être universellement disponible. Il est permis d'utiliser des jeux partiels de polices si tous les glyphes utilisés dans le document sont présents.

Pour les fichiers PDF/A-1a, il est obligatoire d'inclure avec chaque police la correspondance entre les glyphes et les codes Unicode correspondants.

Transparence

Il n'est pas permis d'utiliser les fonctions de transparence dans un fichier PDF/A-1.

Impression

Demi-tons : Il est permis dans un fichier PDF d'indiquer les motifs à utiliser pour les demi-tons. Un fichier PDF/A-1 peut contenir de telles définitions, mais un lecteur n'est pas obligé de les prendre en compte.

Métadonnées

Dictionnaire des informations du document : Le dictionnaire d'information est une façon de stocker les métadonnées de base sur le document. C'était la seule façon de faire dans les versions originales de PDF, mais une deuxième façon (les propriétés de métadonnées XMP) a été ajoutée depuis. Les mêmes informations peuvent être stockées dans les deux façons. PDF/A-1 n'exige pas la présence de ce dictionnaire, mais exige que s'il est présent, son contenu soit cohérent avec les informations dans les propriétés de métadonnées XMP.

Métadonnées XMP : Tout fichier PDF/A-1 doit contenir des métadonnées au format XMP. Ces métadonnées ne doivent pas être compressées pour permettre à des logiciels qui ne comprennent pas le format PDF de pouvoir les lire. Le contenu des métadonnées n'est pas fixe et le créateur du fichier est libre d'utiliser les extensions nécessaires pour bien identifier son document selon les besoins du document.

Les métadonnées d'historique sont préconisées pour garder une liste exhaustive de l'historique du document.

Pour qu'un fichier soit identifié en étant conforme PDF/A-1, il doit contenir des métadonnées XMP qui identifient le niveau de conformité.

Structure logique de document

Un fichier PDF/A-1a est obligé de définir la structure logique du document. Un fichier PDF/A-1b n'en est pas obligé.

Le but de la structure logique est de permettre la récupération du contenu textuel du document correctement et dans le bon ordre. Un fichier PDF de base n'est pas obligé de faire autrement que de définir le positionnement de traits sur une page — même si ces traits forment des mots rien ne dit que le texte peut être récupéré. Avec la définition de la structure logique, cette récupération devient possible.

Guide Méthodologique

Le format PDF

La structure logique est définie par l'utilisation de la fonctionnalité PDF Balisé. Il permet de définir le texte avec une bonne définition des fins de mots (pas toujours facile à récupérer du PDF standard). Il définit aussi la fonction de chaque paragraphe (titre, note de bas de page, titre courant etc), ainsi que la langue utilisée.

Une description textuelle est aussi à fournir pour les éléments graphiques (images, formules etc.), ainsi que l'expansion du texte pour les sigles et acronymes.

Contenu facultatif

Il est possible dans un fichier PDF de créer des contenus qui peuvent ou non être affichés, permettant ainsi de présenter un document de plusieurs façons différentes. PDF/A-1 interdit de tels contenus, obligeant ainsi un document à avoir une seule visualisation possible.

Annotations

Les annotations sont permises dans un fichier PDF/A-1 mais doivent toujours être définies comme visibles et imprimables, sans transparence. Ces critères permettent d'assurer que le document imprimé sera identique à celui vu à l'écran.

Liens hypertexte

Il est possible de mettre des liens hypertexte dans un document PDF/A-1, mais il est mentionné qu'un lecteur pourrait les rendre inactifs. Ainsi même s'il y a des liens hypertexte dans un document, il n'est pas garanti qu'ils soient utilisables.

Compression

Les méthodes de compression permises par PDF 1.4 sont permises dans un fichier PDF/A-1, sauf la compression LZW (qui contient des éléments brevetés).

Sécurité

Cryptage : Il est interdit d'utiliser les fonctions de cryptage et de protection par mot de passe de PDF dans un fichier PDF/A-1. Son utilisation rendrait le fichier difficilement récupérable à long terme.

Formulaires interactifs « AcroForms »

Un fichier PDF/A-1 peut contenir des formulaires interactifs « AcroForms ». Ces formulaires doivent être décrits de façon à ce que leur apparence soit complètement définie.

Multimédia

Il n'est pas permis de mettre des contenus audio ou vidéo dans un fichier PDF/A-1.

Programmation

Le format PDF de base permet de définir des actions et d'inclure des programmes dans un fichier PDF. Toute programmation est interdite dans un fichier PDF/A-1 pour éviter des résultats imprévisibles. Les seules actions permises sont les quatre actions qui permettent de changer de page (première page, dernière page, page suivante, page précédente).

Fichiers embarqués

PDF permet l'inclusion d'autres fichiers à l'intérieur du fichier PDF (un peu comme un fichier Zip contient d'autres fichiers). Cette fonctionnalité est interdite par PDF/A-1, puisque il serait impossible de garantir la lisibilité des fichiers embarqués (qui pourrait être de n'importe quel type) à long terme.

Guide Méthodologique

Le format PDF

OPI

L'utilisation d'OPI pour remplacer des images basse résolution par la haute résolution est interdite.

PDF/A-2

Numéro de norme	ISO 19005-2
Date de publication	1 ^{er} juillet 2011
Titre	Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme — Partie 2: Utilisation de l'ISO 32000-1 (PDF/A-2)
Version de base de PDF	ISO 32000-1 (1.7)
Identifiant PRONOM	—

L'objectif de cette norme est de :

- utiliser une version normalisée de PDF comme version de base (ISO 32000-1) ;
- fournir un format d'archivage plus riche que PDF/A-1 avec les nouveautés qui ont été introduites depuis PDF 1.4.

Il est important de comprendre que cette norme **ne remplace pas ISO 19005-1**, et ne le remplacera jamais. Les deux normes continueront à coexister. Ainsi tout fichier créé selon PDF/A-1 sera toujours considéré comme un fichier valable pour l'archivage. Il ne sera jamais nécessaire de convertir les fichiers PDF/A-1 en PDF/A-2.

La description de la norme ici indiquera tout ce qui est dans PDF/A-2 et mettra en exergue les différences par rapport à PDF/A-1. Il sera donc possible de lire cette section sans avoir retenu tout le contenu de PDF/A-1.

Les buts proclamés de cette norme sont identiques aux buts de PDF/A-1, c'est-à-dire :

- Définir un format de fichier basé sur le PDF pour la représentation de documents dans une manière à préserver leur apparence visuelle dans le temps, indépendamment des outils utilisés pour les créer, les stocker ou les visualiser ;
- Fournir un cadre pour l'enregistrement du contexte et de l'historique de ces documents avec l'utilisation de métadonnées à l'intérieur des fichiers ;
- Définir un cadre pour la représentation de la structure logique et d'autres informations sémantiques à l'intérieur de ces fichiers.

La norme décrit seulement le format des fichiers et ne couvre pas les autres aspects d'archivage tel que le stockage des fichiers ou la validation que le fichier créé est bien une copie conforme du document originel.

L'introduction de la norme liste cinq domaines principaux où il y a amélioration par rapport à PDF/A-1 :

- amélioration du PDF Balisé ;
- meilleures possibilités de compression des contenus du fichier PDF (pour des fichiers plus petits) ;
- possibilité d'embarquer d'autres fichiers PDF/A à l'intérieur d'un fichier ;
- utilisation de la transparence ;
- utilisation de la compression JPEG 2000.

On peut aussi mentionner :

- utilisation de signatures numériques

Niveau de conformité

La norme définit trois niveaux de conformité qu'un fichier peut avoir par rapport à la norme :

- Niveau A (avancé) — le fichier est conforme à toutes les exigences de la norme. Un fichier avec ce niveau de conformité est identifié comme étant du PDF/A-2a ;
- Niveau B (basic) — le fichier est conforme à toutes les exigences de la norme par rapport à l'aspect visuel du document, mais ne contient pas les informations de contenu textuel ou de structure logique du document. Un fichier avec ce niveau de conformité est identifié comme étant du PDF/A-2b ;

- Niveau U (Unicode) — le fichier est conforme à toutes les exigences de la norme par rapport à l'aspect visuel du document et aussi que le texte doit être récupérable en Unicode, mais ne contient pas les informations de structure logique du document. Un fichier avec ce niveau de conformité est identifié comme étant du PDF/A-2u.

Différence avec PDF/A-1 : Le niveau U est nouveau avec PDF/A-2 et n'a pas de correspondance avec une fonctionnalité de PDF/A-1.

Structure du fichier

Besoins du document : Le dictionnaire des besoins du document introduit par PDF 1.7 est interdit dans un fichier PDF/A-2. Puisque les besoins d'un fichier PDF/A-2 sont strictement contrôlés par cette norme, l'inclusion de ces informations serait superflue.

Linéarisation : La linéarisation d'un fichier PDF/A-2 est permise, mais pas obligatoire. Un lecteur conforme PDF/A-2 est censé ignorer les informations de linéarisation.

Dépendances externes : Il est possible avec PDF de définir des dépendances du fichier PDF avec des fichiers externes. Ces dépendances sont interdites par PDF/A-2 pour que le fichier PDF/A-2 soit complètement autonome.

Limites dans l'implémentation : La spécification de PDF indique un certain nombre de limites dans les tailles et les quantités des éléments qui sont contenus dans un fichier PDF. Dans cette spécification ces limites sont données à titre indicatif par rapport aux valeurs utilisées par Acrobat, mais le respect de ces limites n'est pas exigé. PDF/A-2 rend obligatoire le respect de ces limites.

Modélisation de l'image

Images : Chaque image dans un fichier PDF/A-2 doit avoir une seule représentation (il est possible avec PDF d'en spécifier plusieurs) et il n'est pas permis de laisser au lecteur le soin d'interpoler les images pour améliorer leur rendu (puisque cela aurait l'effet de rendre l'image de façon différente selon l'algorithme utilisé par chaque lecteur).

L'utilisation de la compression JPEG2000 est permise, mais uniquement avec les fonctionnalités de base définies par ISO/IEC 15444-2:2004.

Différence avec PDF/A-1 : La compression JPEG 2000 n'était pas permise avec PDF/A-1.

PostScript : Il est interdit d'inclure dans un fichier PDF/A-2 des instructions PostScript. C'était permis dans les premières versions de PDF, mais fortement déconseillé depuis le 1.4.

Couleur

Espaces de couleur : Dans un fichier PDF/A-2 toutes les couleurs doivent être définies de façon indépendante de la sortie. Si des couleurs dépendantes de la sortie sont utilisées, une intention de sortie doit être fournie pour permettre leur interprétation correcte.

Pour toute couleur qui se base sur un profil ICC, le profil associé doit être embarqué dans le fichier PDF.

Différence avec PDF/A-1 : PDF/A-1 ne permettait pas l'utilisation des couleurs dépendantes de la sortie RVB et CMJN dans le même document. PDF/A-2 la permet si les profils ICC adéquats sont fournis.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/A-2 sauf dans les cas où leur utilisation est nécessaire pour une autre fonctionnalité PDF (certains types de demi-ton par exemple).

Différence avec PDF/A-1 : PDF/A-1 interdisait complètement les fonctions de transfert. Vu que dans ISO 32000-1 leur utilisation est parfois obligatoire, PDF/A-2 permet leur utilisation dans ces cas.

Polices

Les polices utilisées dans un document PDF/A-2 doivent être conforme à ISO 32000-1. Toute police utilisée dans le document doit être embarquée dans le fichier PDF/A-2. Il est interdit d'utiliser des polices dont la licence ne permet pas de les embarquer de façon à être universellement disponible. Il est permis d'utiliser des jeux partiels de polices si tous les glyphes utilisés dans le document sont présents.

Pour les fichiers PDF/A-2a et PDF/A-2u, il est obligatoire d'inclure avec chaque police la correspondance entre les glyphes et les codes Unicode correspondants.

Transparence

Il est permis d'utiliser les fonctions de transparence dans un fichier PDF/A-2.

Différence avec PDF/A-1 : la transparence n'était pas permise dans PDF/A-1.

Impression

Demi-tons : Il est permis dans un fichier PDF d'indiquer les motifs à utiliser pour les demi-tons mais uniquement de façon à ce que le résultat soit toujours identique (ISO 32000 permet des fonctions qui donne un résultat différent selon la résolution de la sortie).

Différence avec PDF/A-1 : Le but est le même mais vu que le PDF de base est différent, PDF/A-2 est obligé de décrire différemment la façon de rendre les demi-tons.

Métadonnées

Dictionnaire des informations du document : Le dictionnaire d'information est une façon de stocker les métadonnées de base sur le document. C'était la seule façon de faire dans les versions originales de PDF, mais une deuxième façon (les propriétés de métadonnées XMP) a été ajoutée depuis. Les mêmes informations peuvent être stockées dans les deux façons. PDF/A-1 n'exige pas la présence de ce dictionnaire, et précise qu'un lecteur conforme ignorera ce dictionnaire.

Différence avec PDF/A-1 : PDF/A-1 précisait que le contenu de ce dictionnaire soit identique aux métadonnées XMP, PDF/A-2 précise que le contenu doit être ignoré.

Métadonnées XMP : Tout fichier PDF/A-2 doit contenir des métadonnées au format XMP. Le contenu des métadonnées n'est pas fixe et le créateur du fichier est libre d'utiliser les extensions nécessaires pour bien identifier son document selon les besoins du document.

Les métadonnées d'historique sont préconisées pour garder une liste exhaustive de l'historique du document.

Pour qu'un fichier soit identifié en étant conforme PDF/A-2 il doit contenir des métadonnées XMP qui identifie le niveau de conformité.

Différence avec PDF/A-1 : les métadonnées peuvent être compressées.

Signatures numériques

Les signatures numériques sont permises dans un fichier PDF/A-2.

Différence avec PDF/A-1 : les signatures numériques n'étaient pas possibles dans les fichiers PDF/A-1.

Présentations

Présentation alternative : Un fichier PDF de base peut définir un diaporama des images dans le fichier. Cette fonctionnalité est interdite dans un fichier PDF/A-2.

Différence avec PDF/A-1 : Bien que cette fonctionnalité fut présente dans PDF 1.4 (mais pas documenté avant PDF 1.5), cette interdiction n'était pas présente dans PDF/A-1.

Structure logique de document

Un fichier PDF/A-2a est obligé de définir la structure logique du document. Un fichier PDF/A-2b ou PDF/A-2u n'y est pas obligé.

Le but de la structure logique est de permettre la récupération du contenu textuel du document correctement et dans le bon ordre. Un fichier PDF de base n'est pas obligé de faire plus que de définir le positionnement de traits sur une page — même si ces traits forment des mots rien ne dit que le texte peut être récupéré. Avec la définition de structure logique cette récupération devient possible.

La structure logique est définie par l'utilisation de la fonctionnalité PDF Balisé. Il permet de définir le texte avec une bonne définition des fins de mots (pas toujours facile à récupérer du PDF standard). Il définit aussi la fonction de chaque paragraphe (titre, note de bas de page, titre courant etc.), ainsi que la langue utilisée.

Une description textuelle est aussi à fournir pour les éléments graphiques (images, formules etc.), ainsi que l'expansion du texte pour les sigles et acronymes.

Contenu facultatif

Il est possible dans un fichier PDF de créer des contenus qui peuvent ou non être affichés, permettant ainsi à un document d'être présenté de plusieurs façons différentes. PDF/A-2 permet l'utilisation de cette fonctionnalité (qui est typiquement utilisé pour des documents multilingues, des versions de document par région, etc.). L'utilisation de facteurs externes (langue du système d'exploitation par exemple) pour déterminer le contenu à afficher est interdite.

Différence avec PDF/A-1 : Dans PDF/A-1 cette fonctionnalité était interdite, elle l'est avec PDF/A-2.

Annotations

Les annotations sont permises dans un fichier PDF/A-2 mais doivent toujours être définies comme visibles et imprimables, sans transparence. Ces critères permettent d'assurer que le document imprimé sera identique à celui vu à l'écran.

Liens hypertexte

Il est possible de mettre des liens hypertexte dans un document PDF/A-1 et un lecteur est censé les afficher.

Différence avec PDF/A-1 : Dans PDF/A-1 le texte préconise plutôt de cacher les liens hypertexte. Avec PDF/A-2 les liens hypertexte sont à mettre en valeur.

Compression

Les méthodes de compression permises par ISO-32000 sont permises dans un fichier PDF/A-2, sauf la compression LZW (qui contient des éléments brevetés).

Sécurité

Cryptage : Il est interdit d'utiliser les fonctions de cryptage et de protection par mot de passe de PDF dans un fichier PDF/A-1. Son utilisation rendrait le fichier difficilement récupérable à long terme.

Formulaires interactifs « AcroForms »

Un fichier PDF/A-2 peut contenir des formulaires interactifs « AcroForms ». Ces formulaires doivent être décrits de façon à ce que leur apparence soit complètement définie.

Formulaires interactifs XFA

Les formulaires de type XFA définis à partir de PDF 1.5 ne sont pas permis dans les documents PDF/A-2.

Guide Méthodologique

Le format PDF

Multimédia

Il n'est pas permis de mettre des contenus audio ou vidéo dans un fichier PDF/A-2.

Programmation

Le format PDF de base permet de définir des actions et d'inclure des programmes dans un fichier PDF. Toute programmation est interdite dans un fichier PDF/A-2 pour éviter des résultats imprévisibles. Les seules actions permises sont les quatre actions qui permettent de changer de page (première page, dernière page, page suivante, page précédente).

Fichiers embarqués

PDF permet l'inclusion d'autres fichiers à l'intérieur du fichier PDF (un peu comme un fichier Zip contient d'autres fichiers). Cette fonctionnalité est limitée par PDF/A-2 à l'inclusion de fichiers étant eux-mêmes conformes à une des normes PDF/A.

Différence avec PDF/A-1 : PDF/A-1 ne permettait aucune inclusion d'autres fichiers. PDF/A-2 permet l'inclusion d'autres fichiers qui sont eux-mêmes conformes PDF/A.

OPI

L'utilisation d'OPI pour remplacer des images basse résolution par la haute résolution est interdite.

PDF/A-3

Numéro de norme	ISO 19005-3 (future)
Date de publication	(future)
Titre	Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme — Partie 3 : Utilisation de l'ISO 32000-1 avec support de fichiers incorporés (PDF/A-3)
Version de base de PDF	ISO 32000-1 (1.7)
Identifiant PRONOM	—

Cette norme n'est pas encore publiée et le texte final n'est pas arrêté. Les informations données ici sont donc de nature à pouvoir être modifiées avant sa publication.

Cette norme est identique à la norme PDF/A-2 en tous ses aspects. La seule différence est dans les possibilités d'incorporation des fichiers embarqués.

Dans PDF/A-2 ces fichiers ne peuvent être que des fichiers eux-mêmes conformes aux normes PDF/A. Dans le PDF/A-3 cette restriction est enlevée pour que tout type de fichier puisse être incorporé.

Un mécanisme est défini pour lier le contenu du fichier PDF avec les fichiers embarqués (pour indiquer un fichier source du document, un fichier alternatif pour un tableau etc.). Ce mécanisme n'existe pas dans PDF 1.7 (sur lequel est basée cette norme) mais sera introduit de façon identique dans PDF 2.0 pour que le mécanisme puisse être plus générique que le PDF/A-3 seul.

L'objectif de cette norme est de permettre aux fichiers PDF/A de contenir des collections des documents sources permettant ainsi l'archivage du document résultat avec les fichiers qui ont été utilisés pour le créer.

Il n'est, bien sûr, pas garanti que les fichiers source soient lisibles à long terme, mais ils seront au moins disponibles si nécessaire.

Tableau récapitulatif des normes PDF/A

Fonction	PDF/A-1b	PDF/A-1a	PDF/A-2b PDF/A-3b	PDF/A-2u PDF/A-3u	PDF/A-2a PDF/A-3a
Linéarisation	permis				
Dépendances externes	interdit				
Limites dans l'implémentation	obligatoire				
Besoins du document	—		interdit		
Images	pas de remplacement ou d'interpolation				
Instructions Postscript	interdit				
Motifs de demi-tons	permis, mais à ignorer		permis sauf si ambigu		
Couleur	Indépendant de la sortie, RVB ou CJMN		profil ICC obligatoire si nécessaire pour rendre la couleur indépendante de la sortie		
Profil couleur ICC	permis				
Fonctions de transfert	interdit		interdit, sauf dans les cas les rendant nécessaires		
Polices — glyphes	tous glyphes embarqués				
Polices — correspondance Unicode/glyphes	facultatif	obligatoire	facultatif	obligatoire	
Transparence	interdit		permis		
Métadonnées	obligatoire				
Dictionnaire d'information sur le document	permis (conforme XMP)		permis, mais à ignorer		
Signatures numériques	aucune mention (pas interdit)		permis		
Présentation alternative	ambigu (pas interdit, mais pas documenté dans PDF 1.4 même si présent)		interdit		
Structure logique de document	facultatif	obligatoire	facultatif		obligatoire
Contenu facultatif	interdit		permis		
Annotations	permis, avec restrictions				
Liens hypertexte	permis, mais pas nécessairement utilisables		permis		
Compression	permis sauf LZW				
Compression JPEG 2000	—		permis avec fonctionnalités de base		
Cryptage	interdit				
Protection par mot de passe	interdit				
Formulaires interactifs « AcroForms »	permis si apparence est complètement définie				
Multimédia	interdit				
Programmation	aucune programmation, actions limitées				
Fichiers embarqués	interdit		PDF/A-2 : uniquement des fichiers PDF/A PDF/A-3 : en fichiers associés		

PDF/E

Le but des normes PDF/E est de fournir un format de fichiers pour l'utilisation dans le monde de la documentation d'ingénierie. Plusieurs normes sont prévues pour adresser les formats de fichiers, les workflows et les besoins de transfert de données. Ces normes permettront l'échange et la collaboration de fichiers dans les processus d'ingénierie à la fois à l'intérieur de l'entreprise et entre les différentes entreprises concernées par la documentation technique.

Les fichiers créés dans de tels processus ont souvent plusieurs sources et il est important que ces fichiers soient à la fois fiables et compris de la même manière par chaque entité qui doit les lire et les créer.

PDF/E-1

Numéro de norme	ISO 24517-1
Date de publication	15 mai 2008
Titre	Gestion de documents — Format de documents d'ingénierie utilisant le PDF — Partie 1: Utilisation du PDF 1.6 (PDF/E-1)
Version de base de PDF	1.6
Identifiant PRONOM	—

Cette norme était la première dans la série des normes PDF/E. Aujourd'hui il est considéré que cette norme a été élaborée trop tôt et sans assez de communication. Elle a donc été peu mise en pratique et on attend que la version PDF/E-2 soit bien mieux adaptée aux besoins de l'ingénierie. Vu que le PDF/E-2 avance bien au niveau de la normalisation il est suggéré que, si on n'a pas déjà adopté PDF/E-1, il vaut mieux attendre la sortie de PDF/E-2.

Niveau de conformité

La norme ISO 24517-1 ne définit qu'un niveau de conformité. Un fichier conforme à cette norme serait identifié comme un fichier PDF/E-1.

Structure du fichier

Identifiant : Un fichier PDF/E-1 doit comporter un identifiant unique qui ne doit pas être crypté.

Syntaxe : Plusieurs d'exigences au niveau de la syntaxe du fichier doivent être respectées pour éviter des ambiguïtés :

- l'entête doit commencer au premier octet du fichier ;
- des chaînes hexadécimales doivent toujours comporter un nombre pair de chiffres ;
- la longueur des flux (streams) doit toujours être identique à la longueur déclarée.

Linéarisation : La linéarisation d'un fichier PDF/E-1 est permise, mais pas obligatoire. Un lecteur conforme PDF/E-1 est censé ignorer les informations de linéarisation.

Dépendances externes : Les dépendances externes sont interdites par PDF/E-1 pour que le fichier PDF/E-1 soit complètement autonome.

Limites dans l'implémentation : Tous les fichiers PDF/E-1 doivent respecter des limites indiquées par la norme dans la taille ou le nombre de certains éléments.

Modélisation de l'image

Images : Il n'est pas permis de laisser au lecteur le soin d'interpoler les images pour améliorer leur rendu (puisque cela aurait l'effet de rendre l'image de façon différente selon l'algorithme utilisé par chaque lecteur).

PostScript : Il est interdit d'inclure dans un fichier PDF/E-1 des instructions PostScript.

Couleur

Intention de sortie : Un fichier PDF/E-1 peut comporter une ou plusieurs intentions de sortie, mais ce n'est pas obligatoire sauf si des couleurs non-calibrées sont utilisées dans le fichier.

Espaces de couleur : Un fichier PDF/E-1 peut utiliser n'importe quel espace de couleurs (sauf DeviceN de type NChannel) mais toutes les couleurs doivent être définies de façon indépendante de la sortie. Si des couleurs dépendantes de la sortie sont utilisées, une intention de sortie doit être fournie pour permettre leur interprétation correcte, et on ne peut pas utiliser des couleurs RVB et CMJN dans le même document.

Pour toute couleur qui se base sur un profil ICC, le profil associé doit être embarqué dans le fichier PDF.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/E-1.

Polices

Toute police utilisée dans le document doit être embarquée dans le fichier PDF/E-1. Il est interdit d'utiliser des polices dont la licence ne permet pas de les embarquer de façon à être universellement disponible. Il est permis d'utiliser des jeux partiels de polices si tous les glyphes utilisés dans le document sont présents.

Impression

Si un fichier précise des limites dans la mise à l'échelle pour l'impression, ces limites doivent être respectées. Il est ainsi possible de créer des documents qui seront toujours imprimés à la bonne échelle.

Métadonnées

Un fichier PDF/E-1 contiendrait des métadonnées qui indiquent ce niveau de conformité. Les métadonnées doivent être en format XMP. Des métadonnées de format dictionnaire des informations du document peuvent être fournies, mais dans ce cas elles doivent être cohérentes avec les métadonnées XMP.

Pour l'utilisation des métadonnées référence est faite à ISO 19005-1.

Signatures numériques

Un fichier PDF/E-1 peut contenir des signatures numériques, mais uniquement les types dont la documentation complète est disponible.

Présentations

Les présentations sont permises dans un fichier PDF/E-1, mais si un lecteur les utilise il doit fournir un mode « présentation » distinct.

Les présentations alternatives ne sont pas permises dans un fichier PDF/E-1.

Structure logique du document

L'utilisation de la structure logique du document n'est pas obligatoire dans un fichier PDF/E-1. Si on veut créer des fichiers avec une fonctionnalité complète de recherche, référence est faite à la norme PDF/A-1a pour les détails de comment faire.

Contenu facultatif

Un fichier PDF/E-1 peut comporter du contenu facultatif.

Annotations

Les annotations sont permises dans un fichier PDF/E-1, mais toute annotation doit être visible à l'écran et à l'impression.

Liens hypertexte

Un fichier PDF/E-1 peut comporter des liens hypertexte, mais un lecteur n'est pas obligé de les rendre actifs.

Sécurité

Cryptage : un fichier PDF/E-1 peut être crypté.

Formulaires interactifs « AcroForms »

Les formulaires interactifs « AcroForms » sont permis dans un fichier PDF/E-1.

Formulaires interactifs XFA

Les formulaires interactifs XFA ne sont pas permis dans un fichier PDF/E-1.

Multimédia

Les éléments multimédia autonomes sont permis dans un fichier PDF/E-1 s'ils sont contenus entièrement dans le fichier PDF.

Illustrations 3D

Une illustration 3D doit être de type U3D et ne peut pas dépendre d'actions JavaScript.

Les illustrations 3D sont définies uniquement en couleurs RVB dépendantes de la sortie. Il n'est pas nécessaire de gérer ces couleurs de façon indépendante de la sortie, mais si c'est fait ça doit être fait selon les règles données dans la norme pour la couleur RVB dépendante de la sortie.

Note : cette norme n'assure pas donc le rendu des couleurs utilisées dans les illustrations 3D.

Programmation

Toute programmation est interdite dans un fichier PDF/E-1 pour éviter des résultats imprévisibles.

Fichiers embarqués

Un fichier PDF/E-1 peut inclure des fichiers embarqués.

OPI

L'utilisation d'OPI n'est pas permise dans un fichier PDF/E-1.

PDF/E-2

Numéro de norme	ISO 24517-2
Date de publication	(future)
Titre	Gestion de documents — Format de document d'ingénierie utilisant le PDF — Partie 2 : Utilisation du ISO 32000-2 (PDF/E-2)
Version de base de PDF	ISO 32000-2 (PDF 2.0)
Identifiant PRONOM	—

Cette norme qui est encore en développement vise à remplacer PDF/E-1 avec une norme bien plus complète pour l'ingénierie. Elle sera basée sur PDF 2.0 pour pouvoir inclure les améliorations telles que l'incorporation de la modélisation 3D PRC.

Au début de son développement, PDF/E-21 avait visé l'archivage des documents d'ingénierie (son texte est adapté de PDF/A-2), mais il vise actuellement le simple remplacement de PDF/E-1.

Il apporte des améliorations notamment dans les domaines suivants :

- améliorations dans les illustrations 3D avec l'utilisation de PRC ;
- informations géomatiques en 2D et 3D ;
- meilleure compression et l'utilisation de JPEG2000.

Niveau de conformité

La norme ISO 24517-1 définit deux niveaux de conformité :

- PDF/E-2 — un fichier qui respecte toutes les exigences de la norme et n'est pas crypté ;
- PDF/E-2e — un fichier qui respecte toutes les exigences de la norme et est encryté .

Cette différence a été apportée pour qu'en général les fichiers PDF/E-2 ne soient pas cryptés, mais que, pour les processus dans lesquels le cryptage est nécessaire, ce soit possible.

Exigences de la norme

La norme donne beaucoup d'indications précises sur la façon d'intégrer les différentes fonctionnalités dans un fichier PDF. On ne les reproduit pas ici — souvent ce n'est que pour rendre obligatoire les préconisations de ISO 32000-2.

Le contenu décrit ici est basé sur l'état de développement de la norme en décembre 2011, ce qui pourrait changer avant sa sortie officielle.

Structure du fichier

Syntaxe : Plusieurs exigences au niveau de la syntaxe du fichier doivent être respectées pour éviter des ambiguïtés. Par exemple :

- l'entête doit commencer au premier octet du fichier ;
- des chaînes hexadécimales doivent toujours comporter un nombre pair de chiffres ;
- la longueur des flux (streams) doit toujours être identique à la longueur déclarée.

Ces exigences forcent une structure plus rigide que celle définie par la norme ISO 32000-2 ; cela n'a rien d'exceptionnel du fait que la plupart des fichiers PDF les respectent déjà.

Besoins du document : Le dictionnaire des besoins du document introduit par PDF 1.7 est interdit dans un fichier PDF/E-2. Puisque les besoins d'un fichier PDF/E-2 sont strictement contrôlés par cette norme, l'inclusion de ces informations serait superflue.

Linéarisation : La linéarisation d'un fichier PDF/E-2 est permise, mais pas obligatoire. Un lecteur conforme PDF/E-2 est censé ignorer les informations de linéarisation.

Dépendances externes : Les dépendances externes sont interdites par PDF/E-2 pour que le fichier PDF/E-2 soit complètement autonome.

Limites dans l'implémentation : Tous les fichiers PDF/E-2 doivent respecter des limites indiquées par la norme dans la taille ou le nombre de certains éléments.

Modélisation de l'image

Images : Un fichier PDF/E-2 ne doit pas comporter des images de substitution pour une image. Il n'est pas permis de laisser au lecteur le soin d'interpoler les images pour améliorer leur rendu (puisque cela aurait l'effet d'avoir un rendu différent selon l'algorithme utilisé par chaque lecteur).

PostScript : Il est interdit d'inclure dans un fichier PDF/E-2 des instructions PostScript.

Couleur

Intention de sortie : Un fichier PDF/E-2 peut comporter une ou plusieurs intentions de sortie, mais ce n'est pas obligatoire sauf si des couleurs non-calibrées sont utilisées dans le fichier.

Espaces de couleur : Un fichier PDF/E-2 peut utiliser n'importe quel espace de couleurs mais toutes les couleurs doivent être définies de façon indépendante de la sortie. Si des couleurs dépendantes de la sortie sont utilisées, une intention de sortie doit être fournie pour permettre leur interprétation correcte.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/E-2.

Polices

Toute police utilisée dans le document doit être embarquée dans le fichier PDF/E-2. Il est interdit d'utiliser des polices dont la licence ne permet pas de les embarquer de façon à être universellement disponible. Il est permis d'utiliser des jeux partiels de polices si tous les glyphes utilisés dans le document sont présents.

Transparence

L'utilisation de la transparence est permise dans un fichier PDF/E-2.

Impression

Si un fichier précise des limites dans la mise à l'échelle pour l'impression, ces limites doivent être respectées. Il est ainsi possible de créer des documents qui seront toujours imprimés à la bonne échelle.

Métadonnées

Un fichier PDF/E-2 contiendrait des métadonnées qui indiquent ce niveau de conformité. Les métadonnées doivent être en format XMP. Des métadonnées de format dictionnaire des informations du document peuvent être fournies, mais leur contenu doit être ignoré par un lecteur.

Signatures numériques

Un fichier PDF/E-2 peut contenir des signatures numériques.

Présentations

Les présentations sont permises dans un fichier PDF/E-2, mais un lecteur conforme doit les ignorer.
Les présentations alternatives ne sont pas permises dans un fichier PDF/E-2.

Structure logique du document

L'utilisation de la structure PDF Balisé est préconisée pour un fichier PDF/E-2 pour que le texte soit récupérable. Ce n'est pas obligatoire, uniquement pour les cas où ce n'est pas possible de créer la structure, pour éviter la création d'une structure fictive uniquement pour rendre un document conforme.

Contenu facultatif

Un fichier PDF/E-2 peut comporter du contenu facultatif.

Annotations

Les annotations sont permises dans un fichier PDF/E-2, mais toute annotation doit être visible à l'écran et à l'impression.

Liens hypertexte

Un fichier PDF/E-2 peut comporter des liens hypertexte. Un lecteur PDF/E-2 peut choisir de rendre actif ou non les liens hypertexte externes, mais pour les besoins d'archivage doit toujours fournir un mécanisme pour indiquer quels sont les destinations de ces liens hypertexte.

Compression

Tous les types de compression définis par ISO 32000-2 peuvent être utilisés pour un fichier PDF/E-2 sauf la compression LZW (qui contient des éléments brevetés).

Formulaires interactifs « AcroForms »

Les formulaires interactifs « AcroForms » sont permis dans un fichier PDF/E-2.

Formulaires interactifs XFA

Les formulaires interactifs XFA ne sont pas permis dans un fichier PDF/E-2.

Multimédia

L'utilisation de multimédia (audio ou vidéo) n'est pas permise dans un fichier PDF/E-2.

Illustrations 3D

Une illustration 3D peut être de type U3D ou PRC et ne peut pas dépendre d'actions JavaScript.

Les illustrations 3D sont définies uniquement en couleurs RVB dépendantes de la sortie. Il n'est pas nécessaire de gérer ces couleurs de façon indépendante de la sortie, mais si c'est fait ça doit être fait selon les règles données dans la norme pour la couleur RVB dépendante de la sortie.

Programmation

Toute programmation est interdite dans un fichier PDF/E-2 pour éviter des résultats imprévisibles. Les seules actions permises sont les quatre actions qui permettent de changer de page (première page, dernière page, page suivante, page précédente).

Guide Méthodologique

Le format PDF

Fichiers embarqués

Un fichier PDF/E-2 peut inclure des fichiers embarqués et doit utiliser le mécanisme de fichiers associés introduit dans ISO-32000-2.

OPI

L'utilisation d'OPI n'est pas permise dans un fichier PDF/E-2.

PDF/UA

La norme PDF/UA a comme but de décrire comment créer des fichiers PDF de façon à être accessible par le plus grand nombre de personnes en tenant compte des personnes handicapées d'une façon ou d'une autre.

Les types d'handicap visés sont surtout visuels puisque le PDF est d'abord un format visuel, mais d'autres sont aussi pris en compte :

- cécité/malvoyance — permettre l'utilisation de technologies de lecture audible ;
- daltonisme — interdire l'utilisation de couleur comme seule façon de présenter une information dans un document ;
- surdité — assurer que l'utilisation de son dans un fichier n'est pas la seule façon utilisée pour fournir une information ;
- handicap moteur — assurer qu'un fichier est navigable par d'autres moyens que par la souris seule.

La norme est un mélange de plusieurs types de document :

- format de fichier (les fonctionnalités PDF à utiliser ou non, et comment) ;
- guide de style pour l'écriture de documents ;
- guide des meilleurs pratiques pour la création de documents ;
- spécifications pour les lecteurs et créateurs de fichiers PDF accessibles.

La nature de certaines des consignes données fait en sorte qu'il n'est pas possible de valider automatiquement la conformité PDF/UA d'un fichier PDF. Les éléments de format de fichier peuvent être validés, mais les éléments de style et les pratiques utilisées ne sont pas vérifiables par ordinateur.

PDF/UA-1

Numéro de norme	ISO/DIS 14289-1
Date de publication	En cours
Titre	Applications de la gestion de documents — Amélioration de format du fichier du document électronique pour l'accessibilité — Partie 1: Utilisation de l'ISO 32000-1 (PDF/UA-1)
Version de base de PDF	ISO 32000-1 (1.7)
Identifiant PRONOM	—

Cette norme n'est pas encore publiée, mais est en cours de préparation. Les indications données ici étaient correctes en décembre 2011.

Niveau de conformité

La norme ne définit qu'un seul niveau de conformité : PDF/UA-1 qui couvre la totalité de la norme.

Général

PDF/UA-1 précise quelques règles générales sur la conception de documents :

- interdiction d'utiliser des éléments qui clignotent ;
- interdiction de présenter des informations uniquement par contraste, couleur ou positionnement visuel ;
- toute information présentée par son doit aussi être disponible sans son.

Structure du fichier

Les PDF imbriqués ne sont pas permis dans les fichiers PDF/UA-1.

Les formes composites peuvent être utilisées mais si une forme est utilisée plusieurs fois aucune partie de son contenu ne peut être associée à un élément de la structure logique du document.

Polices

Toutes les polices utilisées dans le document doivent être incluses dans le fichier PDF et elles doivent toutes comporter un codage Unicode. Ces exigences assurent que les glyphes affichés sur les pages correspondent au contenu textuel et donc que le texte disponible pour les technologies d'accessibilité est compris de la même façon que le texte affiché sur la page.

Métadonnées

Le document doit contenir au moins son titre dans les métadonnées.

Sommaire

Le document doit inclure un sommaire qui correspond à l'ordre de lecture du document.

Signatures numériques

L'apparence de la signature doit suivre les règles générales d'accessibilité décrites par PDF/UA. Si une partie d'une signature numérique est représentée par une image, du texte de remplacement doit être fourni pour la décrire.

Articles

S'il y a des articles présents, ils doivent refléter l'ordre logique de lecture du document.

Structure logique de document

Le texte du document doit être présent avec la fonctionnalité de structure logique de document du PDF. Le texte doit être dans l'ordre de lecture logique.

Tout texte qui n'est pas prévu pour la lecture (comme des échantillons de police par exemple) doit être balisé avec une description de sa fonction.

Tout élément graphique doit aussi être balisé avec une description, sauf les éléments qui ne sont présents que pour décoration qui doivent être balisés en tant qu'artefacts.

Les titres dans le texte doivent être balisés en tant que titres avec numérotation des niveaux si nécessaire.

Les tableaux doivent être balisés en tant que tableaux, et les entêtes de colonne et de rangée doivent être bien indiqués.

Les formules mathématiques doivent être identifiées avec un texte alternatif.

Contenu facultatif

Tout état de contenu facultatif doit avoir un nom pour pouvoir identifier la fonction de l'état.

Il est interdit de laisser des automatismes modifier l'état du contenu facultatif ; celui-ci est uniquement modifiable sous le contrôle de l'utilisateur.

Annotations

Les annotations sans texte doivent comporter une description textuelle. L'ordre de sélection des annotations par clavier doit être défini.

Liens hypertexte

Les liens hypertexte doivent comporter une description textuelle.

Les liens hypertexte dont la destination est déterminée par la position sur la page (comme des cartes interactives) sont interdits, sauf si leur fonctionnalité est disponible par une autre méthode.

Sécurité

Cryptage : Si un fichier est crypté, ses droits d'accès doivent permettre l'extraction de texte et d'images. Le cryptage ne doit pas empêcher l'utilisation des technologies d'accessibilité.

Formulaires interactifs « AcroForms »

Pour des formulaires qui ne contiennent pas d'éléments interactifs (typiquement créés pour l'impression ou déjà remplis), il faut fournir les descriptions des éléments, ainsi que leurs rôles et leur état.

Formulaires interactifs XFA

Des formulaires XFA statiques peuvent être utilisés dans un fichier PDF/UA-1, mais pas des formulaires dynamiques.

Multimédia

Des éléments multimédia doivent contenir une description de substitution.

Programmation

Un fichier PDF/UA-1 peut contenir des actions lancées par script, mais si l'action provoque un changement de présentation du document ou de position dans le document, l'action doit être annoncée d'une façon conforme aux technologies d'accessibilité.

Guide Méthodologique

Le format PDF

Fichiers embarqués

Tout fichier embarqué doit être conforme aux normes d'accessibilité selon son type. Le fichier PDF doit contenir une description textuelle du fichier embarqué.

Guide Méthodologique

Le format PDF

PDF/VT

La norme PDF/VT s'adresse aux exigences de l'édition (la production en masse de documents d'entreprise ayant une structure générale fixe et des parties variables en fonction du destinataire). En tant que tel, il s'adresse aux marchés qui utilisaient traditionnellement PPML ou AFP.

ISO 16612-2 (PDF/VT)

Numéro de norme	ISO 16612-2
Date de publication	15 août 2010
Titre	Technologie graphique — Échange de données d'impression variables — Partie 2: Utilisation de PDF/X-4 et PDF/X-5 (PDF/VT-1 et PDF/VT-2)
Version de base de PDF	PDF 1.6
Identifiant PRONOM	—

PDF/VT est conçu pour être neutre par rapport à l'architecture de workflow utilisé, mais a besoin de plus de précisions que le simple fichier PDF pour être mis en œuvre. La norme mentionne le JDF (Job Definition Format) en développement par le CIP4 (*International Cooperation for the Integration of Processes in Prepress, Press, and Postpress Organization*) comme étant une façon de préciser les processus de production de façon indépendante du format graphique.

La norme se base sur PDF/X-4 et PDF/X-5 qui à leur tour se basent sur PDF 1.6.

Une norme précédente (ISO 16612-1) se basait sur l'utilisation de PDF avec PPML, mais les deux normes sont complètement indépendantes et fournissent des solutions différentes.

Pour bien comprendre les notions sous-jacentes à PDF/VT, il est important de saisir la notion de séquence de pages. Dans un flux éditique traditionnel, les pages sont décrites les unes après les autres dans la séquence d'impression voulue, chacune contenant des instructions spécifiques pour choisir le type de papier voulu, si l'impression doit être simple face ou recto/verso, etc.

Un flux PDF/VT par contre est vu comme une base de données avec la définition de chaque page, ou séquence de pages, accessible en accès direct. C'est une description de job (souvent connu comme « job ticket ») qui va préciser l'ordre des pages et les besoins de chaque page.

Le workflow PDF/VT est donc beaucoup moins dépendant du périphérique de sortie à utiliser, et il devient beaucoup plus facile de changer de périphérique à la dernière minute si les besoins de production l'exigent.

Niveau de conformité

La norme définit trois niveaux de conformité qu'un fichier peut avoir par rapport à la norme :

- PDF/VT-1 — pour l'échange complet dans un seul fichier (utilisation de PDF/X-4) ;
- PDF/VT-2 — pour l'échange multi-fichier (utilisation de PDF/X-4p, PDF/5g ou PDF/X-5pg) ;
- PDF/VT-2s — pour l'échange en streaming (traitement du début du fichier avant que la fin ait été reçu ou même créé).

Structure du fichier

Parties de document (DParts) : Un fichier PDF/VT doit définir des parties de document (DParts) pour chaque page, ou séquence de pages qui doivent être traitées ensemble. C'est une structuration du fichier PDF qui n'existe pas dans le PDF 1.6 de base, mais qui est ajouté pour les besoins de PDF/VT. Cette structuration a été reprise dans la norme PDF 2.0.

Formes composites : tout objet graphique qui est appelé plus qu'une fois doit être défini comme une forme composite. Il est ainsi possible de réduire la taille du fichier et optimiser l'impression de ces objets. Le nombre de fois que l'objet est utilisé et son périmètre d'utilisation, est indiqué dans le fichier. Dans PDF/VT-2 ces formes peuvent être des PDF imbriqués externes au fichier d'origine.

Métadonnées

Un fichier PDF/VT doit contenir des métadonnées pour l'identifier comme étant PDF/VT.

Streaming

Le niveau de conformité PDF/VT-2s permet de découper le fichier PDF/VT en morceaux pour l'envoi par streaming. Dans ce cas le fichier est envoyé comme une séquence de packages MIME comme définie par les RFC 2045, RFC 2046, RFC 2047 et RFC 2183.

Les morceaux doivent être envoyés de façon à ce que les références soient toujours vers des morceaux qui ont déjà été envoyés. Typiquement on enverra les packages pour les objets partagés (les ressources comme les polices ou les formes composites) suivi par les parties de documents (DParts) dans l'ordre voulu.

PDF/X

PDF/X est la série de normes créées pour l'échange de documents pour l'imprimerie. Ce sont les premières versions de PDF à être normalisées par l'ISO, et de ce fait il existe plus de révisions que pour les autres normes PDF.

Pour donner une idée de la portée des différentes normes, les précisions suivantes peuvent être apportées :

- PDF/X-1 (parties 1 et 4) — impression de documents complets avec CMJN et tons directs ;
- PDF/X-2 (partie 5) — impression de documents avec éléments externes ;
- PDF/X-3 (parties 3 et 6) — impression de documents complets avec des couleurs calibrées ;
- PDF/X-4 (partie 7) — une norme plus complète couvrant à la fois PDF/X-1 et PDF/X-3 ;
- PDF/X-5 (partie 8) — trois extensions à PDF/X-4.

Voilà un tableau des particularités de chaque version :

Niveau de conformité	Partie de ISO 15930	Échange complet(*)	Gestion de couleurs	Espaces de couleur	Version PDF
PDF/X-1:2001	1	Oui	Non	CMJN	1.3
PDF/X-1a:2001	1	Oui	Non	CMJN	1.3
PDF/X-1:2003	4	Oui	Non	CMJN	1.4
PDF/X-2:2003	5	Non	Oui	Gris, RVB, CMJN	1.4
PDF/X-3:2002	3	Oui	Oui	Gris, RVB, CMJN	1.3
PDF/X-3:2003	6	Oui	Oui	Gris, RVB, CMJN	1.4
PDF/X-4	7	Oui	Oui	Gris, RVB, CMJN	1.6
PDF/X-4p	7	Non	Oui	Gris, RVB, CMJN	1.6
PDF/X-5g	8	Non	Oui	Gris, RVB, CMJN	1.6
PDF/X-5n	8	Non	Oui	n-colorant	1.6
PDF/X-5pg	8	Non	Oui	Gris, RVB, CMJN	1.6
(*) L'échange complet implique que toutes les informations nécessaires pour l'impression sont incluses dans le fichier PDF.					

ISO 15930-1 (PDF/X-1:2001)

Numéro de norme	ISO 15930-1
Date de publication	1 ^{er} décembre 2001
Titre	Technologie graphique — Échange de données numériques de préimpression — Emploi de PDF — Partie 1: Échange complet employant les données CMYK (PDF/X-1 et PDF/X-1a)
Version de base de PDF	1.3
Identifiant PRONOM	fmt/145 PDF/X-1a : fmt/157

PDF/X-1 définit la façon d'utiliser PDF pour l'échange de documents pour l'impression avec une définition des couleurs en CMJN (et des tons directs, bien que pas mentionnés dans le titre).

Niveau de conformité

La norme définit deux niveaux de conformité qu'un fichier peut avoir par rapport à la norme :

- PDF/X-1:2001 — permet tout ce qui est indiqué dans la norme ;
- PDF/X-1a:2001 — permet tout ce qui est indiqué dans la norme sauf l'utilisation d'OPI et le cryptage.

En plus, référence est faite à une norme américaine précédente : *ANSI CGATS.12/1-1999, Graphic technology — Prepress digital data exchange — Use of PDF for composite data — Part 1: Complete exchange (PDF/X-1)*. Cette norme américaine était basée sur PDF 1.2 et peut être référencée comme PDF/X-1:1999 (identifiant PRONOM fmt/144).

Dans les descriptions qui suivent PDF/X-1 se réfère aux deux niveaux de conformité, sauf dans les cas où une différence est indiquée par rapport à PDF/X-1a.

Général

La norme ne permet pas l'utilisation de fichiers pré-séparés (où les séparations sont décrites par des pages PDF différentes).

Modélisation de l'image

PostScript : Un fichier PDF/X-1 ne peut pas contenir des instructions PostScript.

Images de substitution : Un fichier PDF/X-1 peut contenir des images de substitution, mais aucune ne doit être marquée pour l'impression. Ainsi l'image visualisée sera toujours celle imprimée.

Couleur

Espaces de couleur : Les seuls espaces de noms permis dans un fichier PDF/X-1 sont :

- CMJN ;
- niveaux de gris ;
- séparation.

Les espaces de couleur *Separation* et *DeviceN* peuvent être utilisés pour définir des couleurs CMJN, des tons directs ou pour d'autres informations pas liées à la couleur (verniss, découpage etc). Ce sont les noms utilisés qui définissent les couleurs exactes à utiliser et en général les noms doivent être des noms bien reconnus (bien qu'aucune liste ne soit fournie par la norme). Ces espaces de couleur doivent utiliser *DeviceGray* ou *DeviceCMYK* pour définir des couleurs de remplacement.

Guide Méthodologique

Le format PDF

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/X-1 sauf dans les cas où leur utilisation est nécessaire pour une autre fonctionnalité PDF (certains types de demi-ton par exemple).

Polices

Toutes les polices nécessaires pour l'impression du fichier doivent être embarquées dans le fichier PDF et elles doivent être utilisées à l'impression du fichier, sans remplacement par des polices disponibles en local.

Impression

Engraissement : Il est obligatoire dans un fichier PDF/X-1 d'indiquer si l'engraisement a été fait ou non. Il n'est pas permis de créer des fichiers dont ce traitement a été fait partiellement.

Taille de la page : La taille de la page doit être définie par MediaBox et soit TrimBox (préféré) ou ArtBox (mais pas les deux). BleedBox et CropBox sont permis et doivent être cohérents.

Métadonnées

Un fichier PDF/X-1 doit contenir des métadonnées pour l'identifier en tant que PDF/X-1, ainsi que des indications de dates et du titre du document.

Annotations

Si un fichier PDF/X-1 comporte des annotations, elles doivent être hors de la région imprimable de la page.

Compression

La compression des fichiers n'est pas obligatoire. Un fichier PDF/X-1 peut utiliser tous les types de compression permis par PDF 1.3 sauf LZW (qui contient des éléments brevetés).

Sécurité

Un fichier PDF/X-1a ne peut pas être crypté. Un fichier PDF/X-1 peut être crypté mais il ne doit pas contenir un mot de passe utilisateur. Ainsi il doit être possible de toujours l'imprimer sans mot de passe.

Programmation

Un fichier PDF/X-1 ne doit pas contenir des actions ou des scripts.

OPI

PDF/X-1 permet l'utilisation d'OPI, mais PDF/X-1a ne le permet pas.

Lorsqu'OPI est utilisé, tous les fichiers référencés par OPI doivent être inclus comme des fichiers embarqués dans le fichier PDF. Les types de fichiers qui peuvent être référencés sont TIFF/IT, TIFF, EPS et DSC.

Pour les images TIFF/IT et TIFF, uniquement les couleurs CMJN et niveaux de gris peuvent être utilisés.

Pour les images EPS et DSC, uniquement des images rastérisées sont permises, les instructions vectorielles sont interdites sauf pour rogner les images rastérisées. Les espaces de couleur dans ces fichiers sont limités de la même façon que les espaces de couleur dans le fichier PDF.

Des profils ICC ne sont pas permis dans les fichiers appelés par OPI.

ISO 15930-2

Aucune norme n'a été publiée avec le nom ISO 15930-2, mais PDF/X-2 a été publié plus tard comme ISO 15930-5.

ISO 15930-3 (PDF/X-3:2002)

Numéro de norme	ISO 15930-3
Date de publication	1 ^{er} septembre 2002
Titre	Technologie graphique — Échange de données numériques de préimpression — Emploi de PDF — Partie 3 : Échange de fichiers complets aptes à la gestion des couleurs (PDF/X-3)
Version de base de PDF	1.3
Identifiant PRONOM	fnt/158

Comme PDF/X-1, PDF/X-3 est basé sur PDF 1.3. La différence est que PDF/X-1 ne permettait que l'utilisation de couleurs CMJN sans profils de couleur (les niveaux de couleur étaient à interpréter directement par des niveaux de colorants), tandis que PDF/X-3 permet des couleurs calibrées.

Niveau de conformité

La norme ne définit qu'un seul niveau de conformité : PDF/X-3 qui couvre la totalité de la norme.

Exigences de la norme

Puisque les normes PDF/X-1 et PDF/X-3 sont très similaires, on n'indiquera ici que les différences avec la norme ISO-15930-1, sachant que la comparaison est avec PDF/X-1a et non PDF/X-1 (le cryptage et le OPI ne sont pas permis dans un fichier PDF/X-3).

Couleur

Un fichier PDF/X-3 peut utiliser les couleurs décrites par ISO-15930-1 ou des couleurs calibrées à base de profils ICC (CalGray, CalRGB or Lab). Dans tous les cas les couleurs doivent être préparées pour le profil de sortie indiqué à l'intérieur du fichier PDF.

ISO 15930-4 (PDF/X-1a:2003)

Numéro de norme	ISO 15930-4
Date de publication	15 décembre 2003
Titre	Technologie graphique — Échange de données numériques de préimpression utilisant le PDF — Partie 4: Échange complet de données d'impression CMYK et «spot colour» utilisant le PDF 1.4 (PDF/X-1a)
Version de base de PDF	1.4
Identifiant PRONOM	fnt/146

Cette norme est une révision de la norme ISO 15930-1 pour PDF/X-1a pour changer la version de base de PDF en PDF 1.4 (au lieu de PDF 1.3). Le niveau de conformité PDF/X-1 (sans « a ») n'a pas été maintenu.

Le résultat de cette norme est qu'un fichier PDF 1.4 peut être utilisé, mais les améliorations apportées (notamment la transparence, les fichiers PDF importés par référence et la compression JBIG2) ne peuvent pas être utilisées dans le fichier. Ainsi un fichier PDF/X-1a :2003 est fonctionnellement équivalent à un fichier PDF/X-1a :2001.

Général

La norme ne permet pas l'utilisation de fichiers pré-séparés (où les séparations sont décrites par des pages PDF différentes).

Structure du fichier

Importation de fichiers PDF : Un fichier PDF/X-1a ne peut pas contenir des fichiers PDF importés par référence.

Modélisation de l'image

PostScript : Un fichier PDF/X-1a ne peut pas contenir des instructions PostScript.

Images de substitution : Un fichier PDF/X-1a peut contenir des images de substitution, mais aucune ne doit être marquée pour l'impression. Ainsi l'image visualisée sera toujours celle imprimée.

Couleur

Espaces de couleur : Les seuls espaces de noms permis dans un fichier PDF/X-1a sont :

- CMJN ;
- niveaux de gris ;
- séparation.

Les espaces de couleur *Separation* et *DeviceN* peuvent être utilisés pour définir des couleurs CMJN, des tons directs ou pour d'autres informations pas liées à la couleur (vernis, découpage etc). Ce sont les noms utilisés qui définissent les couleurs exactes à utiliser et en général les noms doivent être des noms bien reconnus (bien qu'aucune liste ne soit fournie par la norme). Ces espaces de couleur doivent utiliser *DeviceGray* ou *DeviceCMYK* pour définir des couleurs de remplacement.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/X-1a sauf dans les cas où leur utilisation est nécessaire pour une autre fonctionnalité PDF (certains types de demi-ton par exemple).

Polices

Toutes les polices nécessaires pour l'impression du fichier doivent être embarquées dans le fichier PDF et elles doivent être utilisées à l'impression du fichier, sans remplacement par des polices disponibles en local.

Transparence

Les effets de transparence partiels ou complets sont interdits dans les fichiers PDF/X-1a.

Impression

Engraissement : Il est obligatoire dans un fichier PDF/X-1a d'indiquer si l'engraisement a été fait ou non. Il n'est pas permis de créer des fichiers dont ce traitement a été fait partiellement.

Taille de la page : La taille de la page doit être définie par MediaBox et soit TrimBox (préférée) ou ArtBox (mais pas les deux). BleedBox et CropBox sont permis et doivent être cohérents.

Métadonnées

Un fichier PDF/X-1a doit contenir des métadonnées pour l'identifier comme étant PDF/X-1a ainsi que des indications de dates et du titre du document.

Annotations

Si un fichier PDF/X-1a comporte des annotations elles doivent être hors de la région imprimable de la page.

Compression

La compression des fichiers n'est pas obligatoire. Un fichier PDF/X-1a peut utiliser tous les types de compression permis par PDF 1.4 sauf LZW (qui contient des éléments brevetés) et JBIG2 (qui n'existait pas dans PDF 1.3).

Sécurité

Un fichier PDF/X-1a ne peut pas être crypté.

Programmation

Un fichier PDF/X-1a ne doit pas contenir des actions ou des scripts.

OPI

Un fichier PDF/X-1a ne doit pas utiliser OPI.

ISO 15930-5 (PDF/X-2)

La norme ISO 15930-5 (PDF/X-2) (identifiant PRONOM fmt/147) n'est plus disponible (la norme a été retirée) et donc n'a pas pu être examinée.

C'était une extension de PDF/X-3 pour l'impression de documents qui avaient besoin d'éléments externes pour être imprimés.

ISO 15930-6 (PDF/X-3:2003)

Numéro de norme	ISO 15930-6
Date de publication	15 décembre 2003
Titre	Technologie graphique — Échange de données numériques de préimpression utilisant le PDF — Partie 6 : Échanges complets de données imprimées aptes à la gestion des couleurs utilisant le PDF 1.4 (PDF/X-3)
Version de base de PDF	1.4
Identifiant PRONOM	fmt/148

Cette norme est une révision de la norme ISO 15930-3 pour PDF/X-3 pour changer la version de base de PDF en PDF 1.4 (au lieu de PDF 1.3).

Le résultat de cette norme est qu'un fichier PDF 1.4 peut être utilisé, mais les améliorations apportées (notamment la transparence, les fichiers PDF importés par référence et la compression JBIG2) ne peuvent pas être utilisées dans le fichier. Ainsi un fichier PDF/X-3:2003 est fonctionnellement équivalent à un fichier PDF/X-3:2002.

Exigences de la norme

Puisque les normes PDF/X-1a:2003 et PDF/X-3:2003 sont très similaires, on n'indiquera ici que les différences avec la norme ISO-15930-4.

Couleur

Un fichier PDF/X-3:2003 peut utiliser les couleurs décrites par ISO-15930-4 ou des couleurs calibrées à base de profils ICC (CalGray, CalRGB or Lab). Dans tous les cas les couleurs doivent être préparées pour le profil de sortie indiqué à l'intérieur du fichier PDF.

ISO 15930-7 (PDF/X-4)

Numéro de norme	ISO 15930-7
Date de publication	2008 (deuxième version avec corrections mineurs 15 juillet 2010)
Titre	Technologie graphique — Échange de données numériques de préimpression utilisant le PDF — Partie 7: Échange complet de données d'impression (PDF/X-4) et échange partiel de données d'impression avec une référence de profil externe (PDF/X-4p) utilisant le PDF 1.6
Version de base de PDF	1.6
Identifiant PRONOM	—

Cette version de PDF/X peut être vue comme une version de consolidation. Plusieurs normes ont déjà été publiées sur PDF/X avec plusieurs révisions chacune.

Il restait en gros trois normes :

- PDF/X-1a pour CMYK ;
- PDF/X-2 pour les échanges partiels ;
- PDF/X-3 pour les couleurs calibrées.

Cette norme peut être vue comme étant un remplacement de toutes ces normes. Elle intègre toutes les colorimétries et contient une version pour l'échange complet et une version pour l'échange partielle. En plus il met à jour la version de PDF à PDF 1.6 (au lieu du PDF 1.4 défini dans les normes précédentes). Il a aussi l'avantage de permettre la transparence et le contenu optionnel.

Niveau de conformité

La norme définit deux niveaux de conformité :

- PDF/X-4 pour l'échange complet de données d'impression ;
- PDF/X-4p pour l'échange partiel de données d'impression (le profil ICC est fourni à part).

PDF/X-4 est à préférer s'il n'y a pas d'avantage significatif dans l'utilisation de PDF/X-4p.

Général

La norme ne permet pas l'utilisation de fichiers pré-séparés (où les séparations sont décrites par des pages PDF différentes).

Structure du fichier

Un fichier PDF/X-4 doit contenir tous les éléments nécessaires à l'impression, rien ne peut être importé par référence sauf si la référence est vers un fichier embarqué à l'intérieur du fichier PDF.

Modélisation de l'image

PostScript : Un fichier PDF/X-4 ne peut pas contenir des instructions PostScript.

Images de substitution : Un fichier PDF/X-4 peut contenir des images de substitution, mais aucune ne doit être marquée pour l'impression. Ainsi l'image visualisée sera toujours celle imprimée.

Couleur

Un fichier PDF/X-4 peut utiliser les couleurs spécifiques à un périphérique de sortie ou des couleurs calibrées à base de profils ICC. Dans tous les cas les couleurs doivent être préparées pour le profil de sortie indiqué à l'intérieur du fichier PDF.

L'utilisation de couleurs n-colorant est précisée par la norme PDF/X-5n.

Fonctions de transfert : Les fonctions de transfert permettent de préciser une façon non-linéaire de rendre les différents niveaux de couleur. Ces fonctions sont interdites dans un fichier PDF/X-4 sauf dans les cas où leur utilisation est nécessaire pour une autre fonctionnalité PDF (certains types de demi-ton par exemple).

Polices

Toutes les polices nécessaires pour l'impression du fichier doivent être embarquées dans le fichier PDF et elles doivent être utilisées à l'impression du fichier, sans remplacement par des polices disponibles en local.

Transparence

La transparence est permise dans les fichiers PDF/X-4.

Impression

Engraissement : Il est obligatoire dans un fichier PDF/X-1a d'indiquer si l'engraissement a été fait ou non. Il n'est pas permis de créer des fichiers dont ce traitement a été fait partiellement.

Taille de la page : La taille de la page doit être définie par MediaBox et soit TrimBox (préféré) ou ArtBox (mais pas les deux). BleedBox et CropBox sont permis et doivent être cohérents.

Métadonnées

Un fichier PDF/X-4 doit contenir des métadonnées pour l'identifier comme étant PDF/X-4, ainsi que des indications de dates et du titre du document, comme des métadonnées XMP. Les métadonnées XMP ne doivent pas être compressées.

Si les métadonnées sont aussi présentes comme dictionnaire des informations du document, les valeurs doivent être cohérentes.

Présentations

Présentation alternative : Un fichier PDF de base peut définir un diaporama des images dans le fichier. Cette fonctionnalité est interdite dans un fichier PDF/X-4 puisque le résultat imprimé serait très différent au résultat à l'écran.

Contenu facultatif

Il est permis d'utiliser le contenu facultatif dans un fichier PDF/X-4. Cette fonctionnalité permet, par exemple, d'inclure plusieurs variantes régionales d'un document dans un seul fichier. Dans l'absence d'autres indications ce serait le contenu par défaut qui serait imprimé.

Annotations

Si un fichier PDF/X-4 comporte des annotations elles doivent être hors de la région imprimable de la page.

Compression

La compression des fichiers n'est pas obligatoire. Un fichier PDF/X-4 peut utiliser tous les types de compression permis par PDF 1.6 sauf LZW (qui contient des éléments brevetés).

Guide Méthodologique

Le format PDF

Sécurité

Un fichier PDF/X-4 ne peut pas être crypté.

Formulaires interactifs XFA

Les formulaires interactifs XFA sont interdits dans un fichier PDF/X-4.

Programmation

Un fichier PDF/X-4 ne doit pas contenir des actions ou des scripts.

OPI

Un fichier PDF/X-4 ne doit pas utiliser OPI.

ISO 15930-8 (PDF/X-5)

Numéro de norme	ISO 15930-8
Date de publication	2008 (deuxième version avec corrections mineurs 15 juillet 2010)
Titre	Technologie graphique — Échange de données numériques de préimpression utilisant le PDF — Partie 8 : Échange partiel de données d'impression utilisant le PDF 1.6 (PDF/X-5)
Version de base de PDF	1.6
Identifiant PRONOM	—

PDF/X-5 n'est pas une nouvelle norme, mais un ensemble de trois extensions pour PDF/X-4 qui permettent des flux d'impressions interdits par PDF/X-4. Cette norme a été publiée en même temps que PDF/X-4.

Niveau de conformité

La norme définit deux niveaux de conformité :

- PDF/X-5g — l'utilisation de contenu graphique externe (basé sur PDF/X-4) ;
- PDF/X-5n — l'utilisation de sorties n-colorant (basé sur PDF/X-4p) ;
- PDF/X-5pg — l'utilisation de contenu graphique externe et de profil ICC externe (basé sur PDF/X-4p).

La combinaison de contenu graphique externe et de n-colorant n'est pas couverte par la norme.

Exigences de la norme

Les fichiers PDF/X-5 doivent être conformes complètement aux exigences de ISO 15930-7 (PDF/X-4) sauf :

- des contenus graphiques peuvent être externes (PDF/X-5g et PDF/X-5pg) ;
- les couleurs définies peuvent être pour une sortie n-colorant (PDF/X-5n).

Structure du fichier

Pour les contenus graphiques externes (PDF/X-5g et PDF/X-5pg) on met dans le fichier PDF une forme composite pour indiquer la position de l'élément externe. Cette forme composite peut inclure une image de prévisualisation pour l'élément externe.

Le mécanisme de PDF imbriqué est utilisé pour trouver l'élément externe pour impression (qui doit être conforme à une des normes PDF/X sauf PDF/X-2). Les éléments externes ne peuvent pas contenir du contenu facultatif.

Couleur

Pour PDF/X-5n les couleurs utilisées dans le fichier PDF doivent faire référence à un profil ICC externe de type n-colorant. Le profil ne peut pas être inclus dans le fichier PDF comme fichier embarqué puisque ce cas est exclu par PDF 1.6.

Ce sont les noms des couleurs dans les espaces de couleur DeviceN ou Separation qui fournissent l'indication en ce qui concerne le colorant à utiliser.

Métadonnées

Les métadonnées PDF/X-5 indiquent que le fichier est conforme à un des niveaux de conformité PDF/X-5.

Références

Nom du Document	Version	Localisation du Document
Référentiel Général d'Interopérabilité	2.0	http://references.modernisation.gouv.fr/rgi-interoperabilite
Article 4 de la loi n°2004-575 du 21 juin 2004		http://www.legifrance.gouv.fr/affichTexteArticle.do;jsessionid=F94BCBFF2632F594747AEC7BBE123F76.tpdjo02v_3?idArticle=LEGIARTI000006421544&cidTexte=LEGITEXT000005789847&dateTexte=20081211
Initiative « Open Source »		http://www.opensource.org/
Définition d'un logiciel libre		http://www.gnu.org/philosophy/free-sw.fr.html