

Le PDF, un outil incontournable

Formats de document, standards ISO, archivage à long terme

Tous les formats de document ne sont pas parvenus à s'imposer en tant que standards et tous ne sont pas à la hauteur des promesses affichées. Que peut-on ou que doit-on attendre d'un format ? Quelles sont les principales caractéristiques d'un format et comment savoir qu'un format est « de qualité » ? Où se cachent les risques et comment les éviter ? Le choix d'un format prend tout son sens dans le contexte de l'archivage à long terme. Un archivage de documents dans un format inapproprié pendant plusieurs années ou à un mauvais niveau de qualité peut avoir de graves conséquences.

PDF et les normes ISO

Un format se démarque du lot : PDF, le format natif d'Adobe Acrobat. Dès le début, il a été conçu pour l'échange de documents, indépendamment de la plateforme et du logiciel. Format éprouvé depuis 21 ans, PDF est implémenté par presque tous les fabricants de logiciel qui génèrent des documents électroniques. Ceci dit, la diffusion à très grande échelle de PDF suscite la crainte d'une dépendance par rapport à Adobe. Pour éviter ce conflit, Adobe ainsi que quelques utilisateurs et entreprises industrielles se sont employés à faire reconnaître PDF comme standard industriel et à garantir son développement par les comités d'ISO. Le premier résultat de cette initiative est la publication du standard ISO 32000-1 en 2008. Il se base sur la version PDF 1.7 d'Adobe. Actuellement, ISO travaille à la version PDF 2.0.

La très large gamme de fonctions de PDF ne convient pas pour tous les domaines d'application, sans compter qu'elles ne sont pas nécessaires à tous les utilisateurs. C'est pourquoi ISO a développé un certain nombre de sous-normes en se basant sur le standard PDF en fonction des principales applications.

| Norme | But |
|--------|---|
| PDF/X | Echange de copies d'impression : le résultat d'impression doit être prévisible |
| PDF/A | Archivage à long terme et électronique : voir le texte |
| PDF/E | Ingénierie : modèles 3D interactifs pour les plans de construction et les documents de production |
| PDF/VT | Impression de données et transactionnelle variable : volume important en cache et streaming |
| PDF/UA | Universal Access : accès illimité et aides à l'utilisation, comme Screen-Reader |

Image 1 : les principales sous-normes de PDF

www.pdf-tools.com

Hans Bärfuss est le fondateur et directeur de PDF Tools AG, une société internationale fortement établie dans le secteur du développement et de la distribution de logiciels. Il est délégué de l'Association Suisse de Normalisation au comité ISO de standardisation des formats de documents et des signatures électroniques. Il est également co-fondateur du PDF/A Competence Center, une association qui se consacre à la diffusion et à la promotion de la norme ISO PDF et PDF/A et président du Swiss Chapter. Il participe à de nombreuses présentations dans des conférences et séminaires et publie des articles spécialisés sur les documents numériques.



Les normes PDF/X, PDF/A, PDF/E, PDF/VT et PDF/UA ne sont pas des formats de données à part entière. Elles se basent toutes sur la norme principale PDF 1.7 et définissent les contraintes et limites en fonction de chaque domaine d'application. Ces règles limitent la gamme de fonctions de PDF 1.7 pour définir des sous-ensembles spécifiques (voir l'image 2).

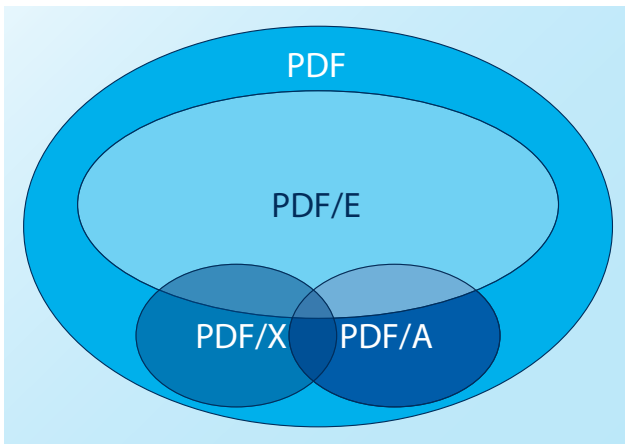


Image 2 : les sous-normes de PDF définissent une partie des fonctions proposées par la norme PDF

PDF pour l'archivage électronique à long terme

Même si PDF a été essentiellement conçu pour l'échange de documents indépendamment des systèmes d'exploitation et du logiciel du créateur et du lecteur du document, ce format ne convient pas nécessairement pour l'archivage. C'est pourquoi PDF/A a été développé.

PDF/A désigne une série de normes qui renferment les caractéristiques de fichiers PDF pour l'archivage électronique à long terme. Ces normes ont pour but de garantir un accès à long terme des documents archivés, de les rendre auto-descriptifs et clairs, de permettre leur description par des métadonnées et de les rechercher. Par conséquent, le cryptage est interdit, les caractères et les profils de couleur doivent être incorporés, les contenus alternatifs et dynamiques doivent être supprimés et les références à des contenus externes disparaître.

En principe, il est également possible qu'un fichier PDF corresponde en même temps à plusieurs sous-normes. Ceci est tout particulièrement important pour l'archivage de fichiers conformes aux normes PDF/X et PDF/VT et qui doivent également être compatibles avec la norme PDF/A. La gamme de fonctions d'un fichier qui revendique la com-

patibilité avec plusieurs normes se limite donc à l'intersection des différentes normes.

Concurrence entre les formats

Bien que très puissant, PDF n'est pas un format magique non plus. Il existe un grand nombre de formats optimisés pour des domaines d'application spécifiques.

| Format de fichier | Conçu pour |
|-------------------|---|
| TIFF | Format d'échange pour les images de quadrillage (scannage, archivage) |
| PostScript/PCL | Langage de description des pages pour l'impression |
| PDF | Echange de documents avec présentation fixe et éléments interactifs |
| AFP | Impression transactionnelle avec données variables |
| Formats Office | Format propriétaire pour les documents de travail dans l'entreprise |
| OOXML | Format Open Office sur XML de Microsoft |
| ODF | Format Open Office en concurrence avec Microsoft |
| XML | Stockage de données orientées objet avec référence vers un schéma |
| XMP | Format de métadonnées extensible sur XML |
| XPS | Langage de description des pages sur XML de Microsoft |
| EPUB | Format pour la publication de documents sur Internet |
| PRC | Stockage de données 3D et d'attributs |
| Multimedia | Nombreux formats pour le streaming audio et vidéo |

Image 3 : formats de fichier et domaine d'application

Dans leurs analyses des avantages et des inconvénients des formats de fichier, de nombreux auteurs ont le réflexe de prendre PDF comme référence pour ensuite établir des comparaisons plus ou moins valables. Cependant, pour que la comparaison tienne la route, il ne faut pas perdre de vue le domaine d'application. Les comparaisons n'ont un sens que lorsqu'il y a recoupement entre les différents domaines d'application. Penchons-nous pour cela sur les domaines qui reviennent le plus souvent dans ces articles.

Boîte de réception

Pour les images de documents scannés, TIFF convient tout autant que PDF. En revanche, les avantages de PDF sont manifestes lorsqu'il s'agit de satisfaire des exigences qui

vont au-delà du simple affichage de pages. D'autre part, les archives TIFF ne doivent pas être converties bêtement en archives PDF en l'absence de valeur ajoutée qui justifierait cette manipulation. Les critères de la conversion d'archives TIFF en archives PDF sont représentés ci-dessus 4.

| | |
|-------------------------------------|----------------------------------|
| <input checked="" type="checkbox"/> | Texte OCR incorporé |
| <input checked="" type="checkbox"/> | Scannage et création numérique |
| <input checked="" type="checkbox"/> | Signature numérique incorporée |
| <input checked="" type="checkbox"/> | Procédés de compression modernes |
| <input checked="" type="checkbox"/> | Commentaires standardisés (XFDF) |
| <input checked="" type="checkbox"/> | Métadonnées incorporées (XMP) |

Image 4 : critères pour la conversion de TIFF à PDF/A

Documents de travail

Les formats les plus souvent utilisés pour les documents de travail sont les formats du pack logiciel Office. OOXML est le format natif de Word, Excel et PowerPoint. Il est impossible de comparer OOXML et PDF selon des critères objectifs. OOXML a été conçu pour la création et l'édition, comprend des informations relatives à la structure et permet la recherche de texte (copier/coller). Il serait bien audacieux de qualifier PDF de format d'auteur ou d'affirmer qu'il convient pour des tâches d'édition. En revanche, en tant que format de présentation fixe et solution d'archivage, PDF s'impose face à OOXML, un format qui n'est absolument pas approprié pour ces tâches, même si certains utilisateurs s'obstinent à penser le contraire.

Cependant, les sous-ensembles XPS et PDF/A présentent de nombreuses caractéristiques communes, comme les contenus statiques, et peuvent donc être comparés. On a justifié le développement de XPS par la nécessité d'éviter toute interruption de format entre le document et le système d'exploitation. Ainsi, le modèle graphique XPS est identique à celui du nouveau sous-système graphique WPF de Vista. Plus précisément, on pourrait dire que les éléments XPS représentent un sous-ensemble de XAML, le langage déclaratif utilisé dans WPF pour les documents et les interfaces utilisateur.

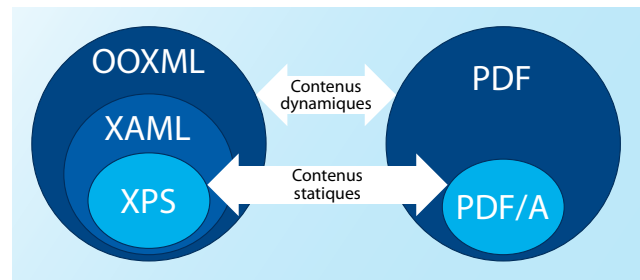


Image 5 : comparaison des formats d'Office avec PDF

Ce cas montre d'ailleurs l'importance de la diffusion d'un format. PDF/A s'est désormais imposé comme le standard pour l'archivage des documents. Au contraire, XPS est presque entièrement inconnu. On pourrait faire le même constat pour ODF, le format Office indépendant de Microsoft.

Boîte d'envoi

Il arrive que les flux de données d'impression au format PostScript, PCL et AFP soient directement archivés. Cependant, la plupart du temps, ils sont convertis au format TIFF (COLD). Au cours des dernières années, on envisageait de passer du tout-tiff à la conversion au format PDF/A. Pour cela, les critères tels que la taille des fichiers (contrainte pour l'intégration des polices), le temps consacré à la conversion, la qualité du résultat, etc. jouent un rôle important. Pour PostScript et PCL, il n'existe aucun standard d'archivage. En revanche, on travaille pour AFP à l'élaboration d'une norme AFP/A comme alternative à PDF/A.

Ces considérations jouent un rôle tout particulièrement important pour qui recherche une solution d'archivage à part entière pour les boîtes d'envoi. Si on donne la priorité à une solution d'archivage unique pour l'ensemble de l'entreprise, on se retrouve rapidement avec le format PDF/A.

Des données en 3D pour les ingénieurs

Pour les données en 3D, qui doivent être manipulées de façon interactive et qui sont enrichies avec des descriptions (par ex. pour les listes de pièces), il existe une norme ISO dénommée PRC (Product Representation Compact). Incorporable dans PDF, PRC est une caractéristique importante de la norme PDF/E. PDF/E-2 a été conçu pour permettre un archivage direct des données compatibles.

Métadonnées

Il existe de nombreux formats propriétaires pour les métadonnées. Souvent, les métadonnées sont également classées directement dans le système d'archivage. Cependant,

il est ici vivement conseillé d'utiliser un format standard. ISO propose pour cela XMP (Extensible Metadata Platform), basé sur XML. XMP peut être incorporé dans tous les formats d'image JPEG, TIFF, etc. et fait partie intégrante de PDF/A. Il est également possible d'incorporer directement des données XML dans PDF/A-3, par exemple pour la facturation électronique (ZUGFeRD).

Multimedia

Les fichiers audio et vidéo sont bien entendu également concernés par l'archivage, soit individuellement, soit sous la forme de flux de données incorporés au format PDF ou autre. Cependant, pour ces formats, aucun standard ne fait encore l'unanimité. On peut espérer que cette situation évoluera à l'avenir. EPUB, un format de publication électronique, joue dans ce contexte un rôle pionnier.

Archivage : PDF/A est-il la solution ?

Tant qu'on reste dans l'univers des documents électroniques, PDF/A reste certainement le format le plus qualifié. Cependant, ce constat ne s'applique pas à toutes les archives. Dans certains domaines, par exemple les moyens de communication interactifs tels que les sites Web, les programmes et les contenus multimédia, on a besoin de nouvelles normes. Les experts ISO ne risquent pas de manquer de travail ces prochaines années.

PDF Tools SA
Kasernenstrasse 1
8184 Bachenbülach
Suisse

Tél: +41 43 411 44 51
Fax: +41 43 411 44 55
pdfsales@pdf-tools.com
www.pdf-tools.com