

PDF/A: validateurs

Comment se définit le PDF/A? Quelles difficultés sont à prendre en compte? Quels validateurs PDF/A courants existe-t-il? Quel est le degré de précision de ces validateurs PDF/A?

1. Introduction

Le format PDF/A est très répandu dans les archives et est un format d'archivage incontesté. Mais il se pose de plus en plus souvent la question de savoir quand un PDF/A est vraiment un PDF/A, quelles ressources permettent de le vérifier et pourquoi les validations ne peuvent pas donner des résultats stricts.

Ce récapitulatif des différents programmes de validation des fichiers PDF/A pourra servir d'aide aux Archives à deux égards: premièrement, il montre quels validateurs existent pour la conversion de fichiers en PDF/A au stade actuel des connaissances; deuxièmement, il peut servir de base à une première évaluation par les différentes Archives.

2. Définition du PDF/A:

Le PDF/A est un format de document destiné à l'échange (Portable Document Format) et qui a été créé pour l'archivage à long terme. Ce format a été standardisé dans la norme «ISO-19005-1 - Document management – Electronic document file format for long-term preservation». Cette norme correspond à la version PDF 1.4. La norme ne fait qu'énumérer quelles fonctions des différentes versions de PDF sont obligatoires, recommandées, restreintes ou interdites.

Le PDF/A existe en deux variantes:

- PDF/A 1a: concordance parfaite avec la norme
- PDF/A 1b: exigences de base de PDF/A satisfaites (critères d'accessibilité non satisfaits)

3. Problèmes généraux pour la validation

Le PDF/A a certes sa norme ISO, mais celle-ci ne fait qu'énumérer quelles fonctions de la version 1.4 de PDF sont obligatoires, recommandées, restreintes ou interdites. Dans les détails, ces fonctions sont interprétées individuellement de manière différenciée. En outre, tous les documents *i* qui définissent conjointement la norme PDF/A sont très volumineux et très techniques. Il s'ensuit que seuls des experts disposant de solides connaissances des langages de description de pages tels que PostScript et PDF peuvent procéder à une évaluation complète sans l'aide de validateurs PDF/A.

Adobe Reader n'affiche pas des résultats suffisamment concluants. Il ne vérifie que quelques caractéristiques principales de PDF/A. Si on ouvre les documents non conformes de la suite des essais Bavaria avec le logiciel Adobe Reader 9.3.3, celui-ci reconnaît seulement 4 des 226 documents non conformes à PDF/A (<2%). Cet affichage peut également faire partie d'Adobe Acrobat et ne correspond pas avec les résultats du Preflight *ii*.

On ne devrait donc pas renoncer à une validation systématique avec des validateurs spécialisés.

i La norme ISO 19005-1, référence PDF 1.4 (env. 1'000 pages), plus les documents qui y sont référencés, tels que formats de polices, spécification XML, formats de compression, RFC etc.

ii Le validateur Preflight est intégré dans Adobe Acrobat et a été développé par callas software GmbH.

4. Evaluation des validateurs PDF/A

Etant donné la problématique mentionnée ci-dessus, il n'est pas facile d'évaluer les validateurs PDF/A existants. Le CECO ne serait pas en mesure de le faire tout seul. C'est pourquoi il s'appuie sur les travaux préliminaires de l'entreprise **PDFlib** *iii*, qui a non seulement fourni une évaluation publique, mais également mis en ligne les fondements contenant ce savoir-faire sous forme du « Bavaria-Testsuite » *iv*.

Avec ces bases et l'accord obtenu de **PDFlib**, le CECO est en mesure d'évaluer d'autres validateurs et de développer l'évaluation existante. En outre, les différents résultats des tests ont été résumés et sommairement évalués, comme on peut le voir dans le tableau ci-après.

5. Commentaires sur le rapport et la suite des essais Bavaria de PDFlib

En tant que fabricant du rapport et de la suite des essais Bavaria, **PDFlib** tient à rendre les utilisateurs attentifs des points suivants:

- Les données d'essai pour le rapport Bavaria ne suffisent pas pour vérifier toutes les violations concevables du standard PDF du à la complexité de ce dernier.
- Il serait certes souhaitable et intéressant d'étendre la suite des essais Bavaria. Cependant, **PDFlib** GmbH n'a pas l'intention d'élargir les tests pour le rapport Bavaria afin d'examiner d'autres éléments de PDF.
- En raison de la disponibilité des données d'essai depuis la publication du rapport Bavaria (avril 2009) les fabricants de logiciels PDF/A pourraient théoriquement traiter les documents PDF du rapport Bavaria de manière séparée et donc se limiter à chercher les documents du test au lieu de chercher tous les éléments critiques. Ceci aboutirait en effet à un bon résultat de test, mais qui ne serait pas forcément suffisamment précis dans la pratique.
- Pour des raisons financières **PDFlib** GmbH n'est pas capable d'assister les utilisateurs pour la suite des essais Bavaria et son application aux validateurs de PDF/A.

6. Nouvelles versions ou autres programmes

Comme les tests sont très coûteux, le CECO renonce à tester séparément chaque nouvelle version. Il ne nous est pas non plus possible de représenter tous les validateurs existants. Le CECO procède cependant volontiers à des actualisations ou à des ajouts sur demande.

iii Les fondements de l'entreprise **PDFlib** sont dignes de confiance, car elle est membre du « **PDF/A Competence Center** » et a en outre besoin de cette évaluation pour ses propres produits (création de documents PDF/A).

iv www.pdfliib.com/knowledge-base/pdfa/validation-report

7. Récapitulatif des produits

Validateurs PDF/A (récapitulatifs de divers programmes desktop)	Adobe: Adobe Acrobat 9.1 www.adobe.com	Callas: pdfaPilot www.callassoftware.com	Intarsys: PDF/A Live www.intarsys.de	PDF Tools: 3Heights PDF Validator Shell www.pdf-tools.com	Seal Systems: PDF Longlife Suite / PDF Checker www.sealsystems.de	Solid Documents: Solid Framework www.soliddocuments.com	PDFTron: PDF/A Manager www.pdftron.com
Vitesse et robustesse: Très bien = <1 minute et sans plantage Bien = 1 à 5 minutes et sans plantage Suffisant = >5 minutes et sans plantage Insuffisant = plantage	Suffisant	Suffisant	Bien	Très bien	Bien	Suffisant	Très bien
Précision: ✓ Très bien = valeur moyenne >=95% Bien = valeur moyenne 90% - 94% Suffisant = valeur moyenne 75 à 89% Insuffisant = valeur moyenne <75%	Bien	Bien	Très bien	Bien	Suffisant	Insuffisant	Très bien
Isartor testsuite (non-conforming)							
6.1 File structure 31x	97%	100%	90%	100%	42%	100%	90%
6.2 Graphics 47x	100%	100%	100%	100%	83%	100%	100%
6.3 Fonts 28x	100%	100%	96%	100%	50%	100%	100%
6.4 Transparency 6x	100%	100%	100%	100%	100%	100%	100%
6.5 Annotations 25x	96%	100%	100%	100%	100%	100%	100%
6.6 Actions 37x	100%	100%	100%	100%	100%	100%	100%
6.7 Metadata 27x	100%	100%	100%	100%	41%	37%	96%
6.9 Interactive Forms 3x	100%	100%	100%	100%	100%	100%	100%
Other non-conforming							
ISO 19005 violations 9x	89%	100%	89%	89%	33%	67%	100%
XMP 2004 violations 5x	20%	40%	60%	20%	40%	40%	80%
PDF 1.4 violations 8x	38%	38%	100%	63%	25%	38%	75%
Conforming							
Real world 34x	88%	100%	97%	85%	82%	41%	100%
PDFlib samples 8x	100%	100%	100%	88%	75%	13%	88%
Advanced XMP 16x	100%	100%	100%	100%	100%	0%	100%
Moyenne (conforming / non-conforming)	91%	94%	97%	90%	75%	49%	95%
Version testée:	9.1.0	1.2.077	5.0.4	1.8.32.1	2.1.1.2	5.1.168	1.00 (CLI)
Testeur / Année du test:	PDFlib / 2009	PDFlib / 2009	CECO / 2010	PDFlib / 2009	PDFlib / 2009	PDFlib / 2009	CECO / 2010
Observations:	La version 9.0 est insuffisante au niveau de la vitesse et de la robustesse.		Test entièrement refait avec une nouvelle version				Le rapport est très clairement structuré (compact avec de bons messages d'erreurs)

PDFA-Validateurs_v2.1.doc

Stand: 08.12.2010

✓ Pour la précision, le résultat (réussi/échoué) n'est pas seulement essentiel; en outre, le message d'erreur doit décrire au moins une erreur réelle.

PDFA-Validateurs_v2.1.doc

Rc, 08.12.2010 Az v0.1

Seite 3/3