



**CENTRE COMMUN
DE LA SÉCURITÉ SOCIALE**

Adresse postale: L-2975 Luxembourg
Bureaux: 125, route d'Esch
Tél.: 40141-1 | Fax: 495764 | www.ccss.lu

DÉPARTEMENT INFORMATIQUE

Spécifications PDF Import externe GED

Historique du document

Version	Date	Auteurs	Modifications effectuées
1.0	02 février 2016	F. Wolter	Version initiale
1.1	09 juin 2016	F. Wolter	Uniformisation pour application à tous les partenaires externes susceptibles de fournir des PDF à archiver
1.2	25 janvier 2017	F. Wolter	Ajout de références vers des logiciels externes
1.3	08 février 2017	F. Wolter	Détails concernant les compressions d'images
1.4	07 mars 2017	F. Wolter	Ajout de détails sur les outils de production PDF/A
1.5	16 mars 2017	F. Wolter	Apis de développement et validateur PDF/A
1.6	12 février 2018	F. Wolter	Configuration d'imprimantes virtuelles

Table des matières

1. OBJET DU DOCUMENT :	3
2. SPÉCIFICATION DU FORMAT PDF	3
3. SPÉCIFICATION DE LA PRODUCTION DES PDF	3
3.1. Cas des documents numérisés	3
3.2. Cas des documents produits électroniquement	4
3.3. Outils permettant la production de PDF/A	4
3.3.1. Outils de développement	4
3.3.2. Imprimante virtuelle	4
3.3.3. Valideur PDF/A	5
4. ANNEXES	5
4.1. Configuration de Solid PDF Creator	5
4.2. Configuration de PDFForge PDFCreator	7
4.3. Installation et utilisation de VeraPDF	7
4.3.1. Installation	7
4.3.2. Utilisation	8
4.4. Création de documents PDF/A	10

1. Objet du document :

Ce document définit les spécifications du format PDF s'appliquant pour la prise en charge par la GED des Institutions de sécurité sociale de documents de source externe.

Ce document est à destination des partenaires externes d'une Institution de sécurité sociale fournissant des documents à archiver dans le système GED d'une Institution de sécurité sociale.

Il contient des informations sur des outils de production et de validation PDF permettant de fournir les documents tels que spécifiés.

2. Spécification du format PDF

- Le format PDF doit être de type PDF/A-1b (ISO ISO 19005-1 Level B) (*)
- Si les PDF sont de type image (aucune fonte ou ressource inutile ne doit être embarquée dans le fichier)
- Les images doivent être compressées (CCITT G4 pour N&B ou JBIG2, JPEG pour 256 niveaux de gris)
- Le format de page PDF à fournir est A4 (exceptionnellement A5 ou A3), en portrait ou en paysage
- De manière optionnelle, uniquement les tags d'information PDF suivants peuvent être renseignés (Author, PDF Producer et Application)

(*) Nous effectuons une validation PDF/A-1b sur tous les documents rentrants. Si le partenaire externe souhaite tester la validité PDF/A de ses documents, plusieurs validateurs PDF/A en ligne gratuits sont disponibles (<http://demo.verapdf.org/>, <http://www.pdf-tools.com/pdf/validate-pdf-a-online.aspx>, <http://www.validatepdfa.com/fr/online.htm>,...)

Un validateur autonome disponible en open source est également utilisable (<http://verapdf.org>)

3. Spécification de la production des PDF

3.1. Cas des documents numérisés

Qualité de la numérisation

- Les documents seront numérisés en 200dpi en noir et blanc
- Les pages blanches ou inutiles (entêtes) seront supprimées
- Exceptionnellement certaines pages peuvent être scannées en 256 niveaux de gris (si l'original contient des photos)

Compression

- Le générateur doit compresser les documents afin de garantir une lisibilité proche de l'image numérisée
 - o En N&B : CCITT G4 garantissant une qualité identique au document numérisé ou JBIG2
 - o En niveaux de gris : jpeg (avec qualité de compression jpeg d'environ 60%)

Contenu du fichier

- Un fichier doit contenir un seul document : aucun retraitement du PDF ne sera effectué

Taille des fichiers :

- Pour des documents numérisés en N&B, on arrive généralement à une taille d'image de 20-40Kb par page.
- Les pages contenant des photos peuvent être scannées en 256 niveaux de gris afin de garantir une lisibilité suffisante. Dans ce cas, le contenu du PDF doit être compressé afin de garantir une taille réduite (150 à 200Kb par page)

Contenu texte :

- Les contenu texte de type OCR plein texte sont acceptés dans la mesure où le texte est caché derrière le niveau image.

3.2. Cas des documents produits électroniquement

Dans le cas de documents produits électroniquement depuis une source texte, le rendu PDF fourni devra respecter les règles suivantes :

- Les zones de texte ne doivent pas être transformées en image (afin de minimiser la taille des documents)
- Les ressources graphiques de type tableau doivent être intégrées sous forme d'objets vectoriel et pas de bitmap.
- Sauf cas spécifiques (à discuter avec le CISS) les documents ne devront pas avoir de bookmark.
- Toutes les fontes utilisées devront être embarquées dans le PDF (exigé par PDF/A-1b)
- Aucune fonte non utilisée ne devra être embarquée dans le PDF.

3.3. Outils permettant la production de PDF/A

Le Département informatique du CCSS ne recommande pas l'utilisation d'un outil particulier permettant de produire du PDF/A, de nombreux outils existant permettent de répondre à ce besoin.

Les 2 chapitres qui suivent donnent des exemple d'outils permettant de produire du PDF/A-1b

3.3.1.Outils de développement

Les APIs PDFlib permettent de produire des fichiers PDF/A-1b et sont utilisables dans de nombreux langages de développement (A noter qu'il s'agit d'une licence payante)

Parmi les APIs gratuites permettant de produire du PDF/A-1b, les plus répandues sont Apache PDFBox et iText.

3.3.2.Imprimante virtuelle

Certaines imprimantes virtuelles permettent de produire des PDF/A-1b valides.

- Parmi les solutions payantes, on peut citer : Adobe et eDocPrintPro PDF/A
- Parmi les solution gratuites, PDFForge PDFCreator et Solid PDF permettent de produire des PDF/A-1b valides.(En annexe procédure de configuration de cet outil).

Il est à noter que de nombreuses autres solutions gratuites prétendent fournir du PDF/A-1b mais le format fourni n'est pas toujours conforme à la norme. (Il convient donc en cas de doute de vérifier à l'aide d'un outil de validation : Voir liste en chapitre 2)

3.3.3. Validateur PDF/A

Un certain nombre d'outil permettent de valider qu'un PDF est bien conforme à la norme PDF/A. Le plus complet est VERAPDF qui est décrit dans ce document en annexe.

Dans le cas d'une intégration d'une intégration de plus bas niveau (sur base d'APIs de développement l'outil Apache PDFBox est aussi utilisable).

4. Annexes

4.1. Configuration de Solid PDF Creator

L'outil décrit dans la suite est un outil gratuit mais non Open-Source, donc dans le cas de l'intégration de cet outil dans une solution commerciale, il incombera à l'intégrateur de s'assurer auprès de l'éditeur qu'il est en règle avec la politique de licence liée au produit.

L'outil est téléchargeable sur le site de l'éditeur à l'adresse suivante : <http://www.soliddocuments.com/en/products.htm?product=SolidPDFCreator>

L'installation peut se faire en utilisant la procédure standard d'installation
Les configurations suivantes doivent être faite

- 1) Sélection de conformité PDF/A-1b

Conformité PDF

Configurer l'option de création du pilote d'imprimante de conformité PDF par défaut.

- ☐ PDF (aucune restriction de conformité)
☒ PDF/A

Le PDF/A est une définition s'appliquant aux fichiers électroniques et assurant l'affichage futur des fichiers dans leur forme véritable. Les fichiers PDF/A doivent être entièrement indépendants et comprendre tout le contenu du fichier (renseignements sur le texte, l'image, les polices et les

Niveau de conformité

PDF/A-1b ▼

En cas de non-conformité

- ☒ Annuler la création PDF
☐ Afficher la mise en garde avec l'option pour continuer

2) Paramètres de qualité des PDF

Paramètres de qualité de création PDF

La taille du fichier PDF dépend de la qualité. Choisissez le paramètre de qualité en fonction de l'utilisation prévue.

☐ Aucun

Compression sans perte, aucun rééchantillonnage d'image, toutes les polices intégrées

☐ Web

Idéal pour l'affichage à l'écran. (Compression avec perte, rééchantillonnage d'image à 150 DPI, polices de sous-ensemble.)

☐ Qualité brouillon

Idéal pour les brouillons ou les tâches d'impression de faible résolution. (Compression avec perte, rééchantillonnage d'image à 150 DPI, aucune police)

☐ Qualité professionnelle

Idéal pour les tâches d'impression de grande qualité. (Compression sans perte, rééchantillonnage d'image à 600 DPI, polices de sous-ensemble.)

☒ Personnalisé

Paramètres

Sélectionnez vos propres paramètres d'optimisation.

Sélectionnez Personnalisé, en cliquant sur paramètres vous obtiendrez l'écran suivant :

Solid PDF Creator - Optimiser PDF

Optimiseur d'image
Optimiseur de police

Images couleur

☒ Convertir les images couleur en échelle de gris

Sous-échantillonnage (DPI) : 150

Compression : Automatique

Qualité de l'image : Moyenne

Images en échelle de gris

☒ Convertir les images en échelle de gris en images monochromes

Sous-échantillonnage (DPI) : 150

Compression : Automatique

Qualité de l'image : Moyenne

Images monochromes

Sous-échantillonnage (DPI) : 150

Compression : CCITTFaxEncode, Group4

OK Annuler

Ces réglages permettent d'obtenir des images en noir et blanc d'une taille réduite et sont valables dans le cas d'images numérisées.

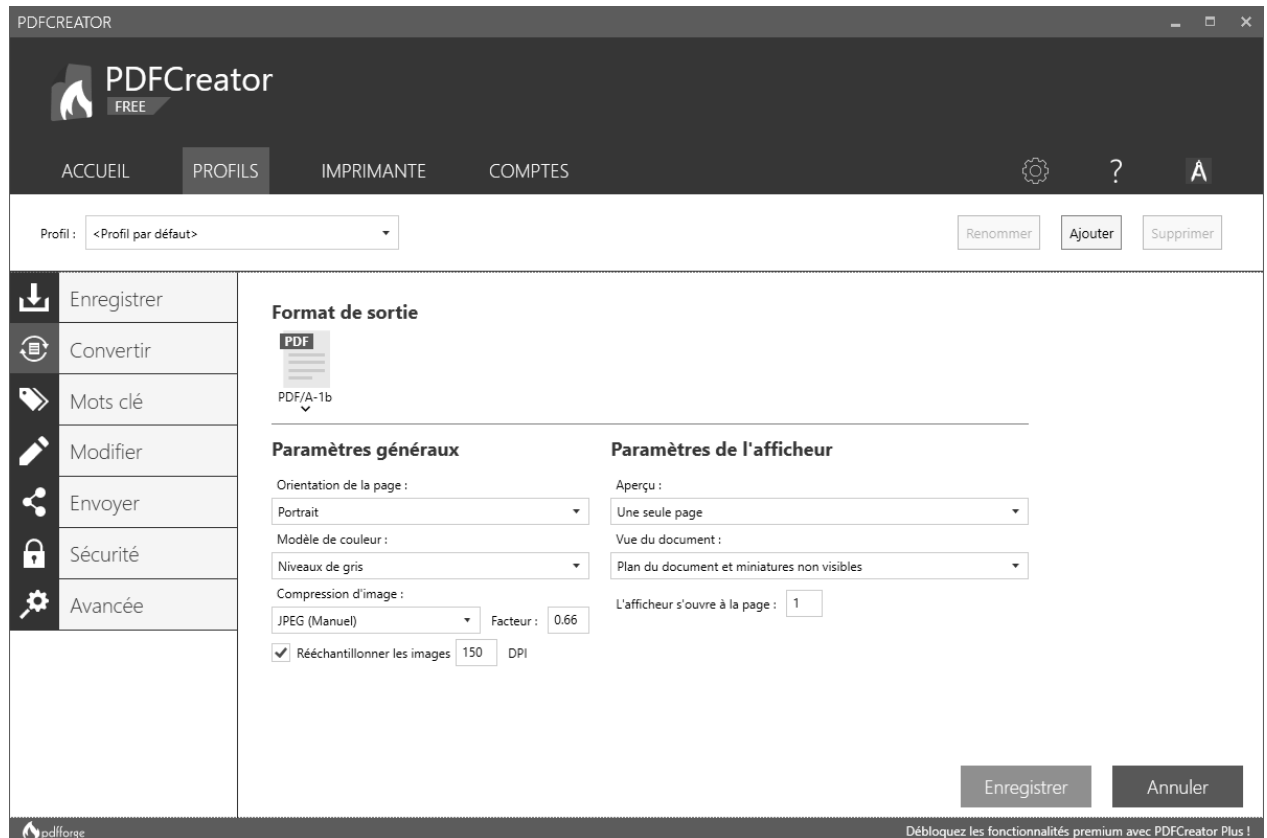
Les réglages pourront être adaptés dans le cas où certaines images doivent être produites en niveau de gris.

4.2. Configuration de PDFForge PDFCreator

L'outil PDFCreator fournit des PDF/A 1-b valides à partir de sa version 3.1.

Il est téléchargeable et installable depuis le site <https://www.pdfcreator.org>

Lors de la configuration, vous devez vous assurer dans la configuration de vos profils de sélectionner le format PDF/A 1-b



4.3. Installation et utilisation de VeraPDF

VeraPDF est un outil de validation PDF/A. Il permet simplement de s'assurer qu'un PDF est bien conforme à la norme PDF/A.

4.3.1. Installation

L'outil est disponible en téléchargement sur <http://verapdf.org/software/>

Sur la même page les instructions d'installation sont fournies.

Dans le répertoire d'installation vous allez obtenir une structure de fichier de ce type :

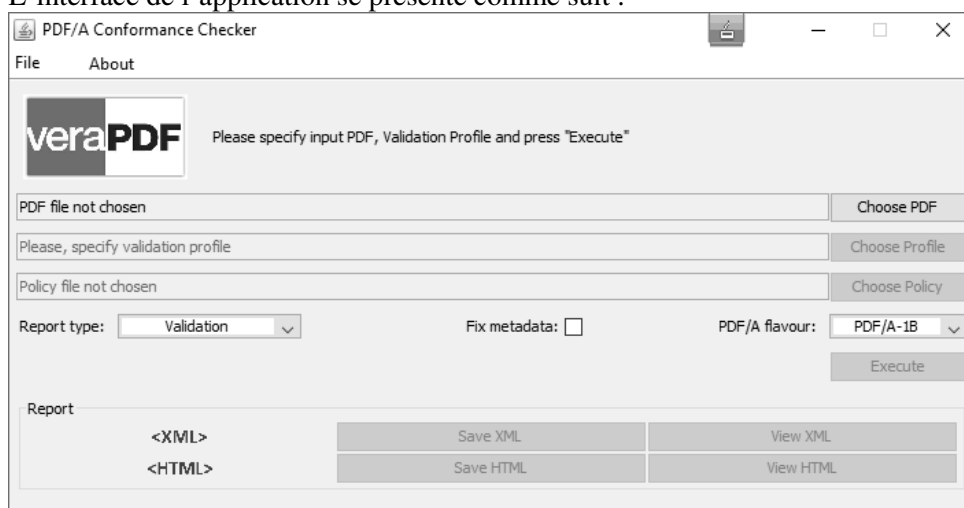
Name	Date modified
bin	14.03.2017 13:54
config	10.02.2017 11:41
corpus	10.02.2017 11:41
documents	10.02.2017 11:41
model	10.02.2017 11:41
plugins	10.02.2017 11:41
profiles	10.02.2017 11:41
Uninstaller	10.02.2017 11:41
.installationinformation	14.03.2017 13:54
verapdf.bat	01.03.2017 22:59
verapdf-gui.bat	01.03.2017 22:59

L'outil verapdf.bat est un outil en ligne de commande pour utilisateurs avancés (voir documentation sur le site verapdf)

L'outil verapdf-gui.bat est un outil graphique permettant la validation d'un PDF. Le paragraphe suivant détaille l'utilisation de ce module graphique.

4.3.2.Utilisation

L'interface de l'application se présente comme suit :

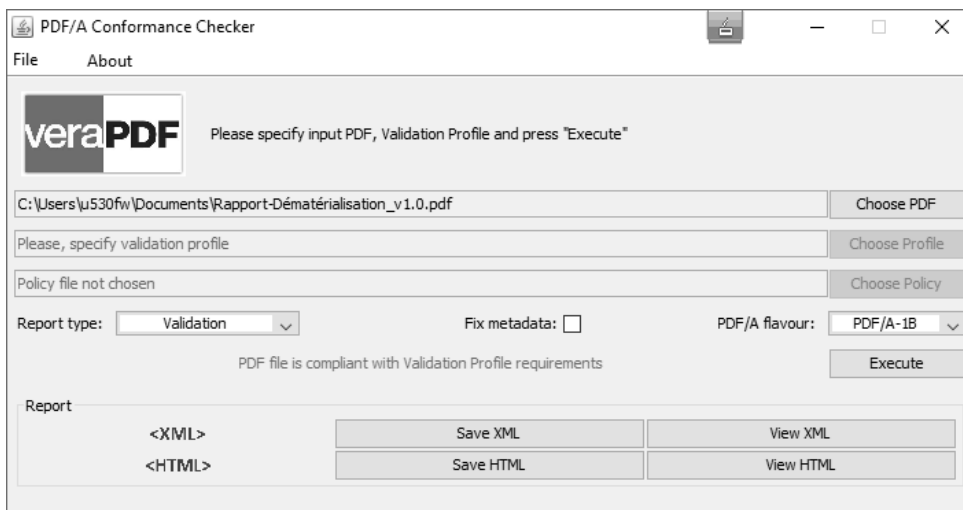


Pour valider qu'un document est un PDF/A valide, procédez comme suit :

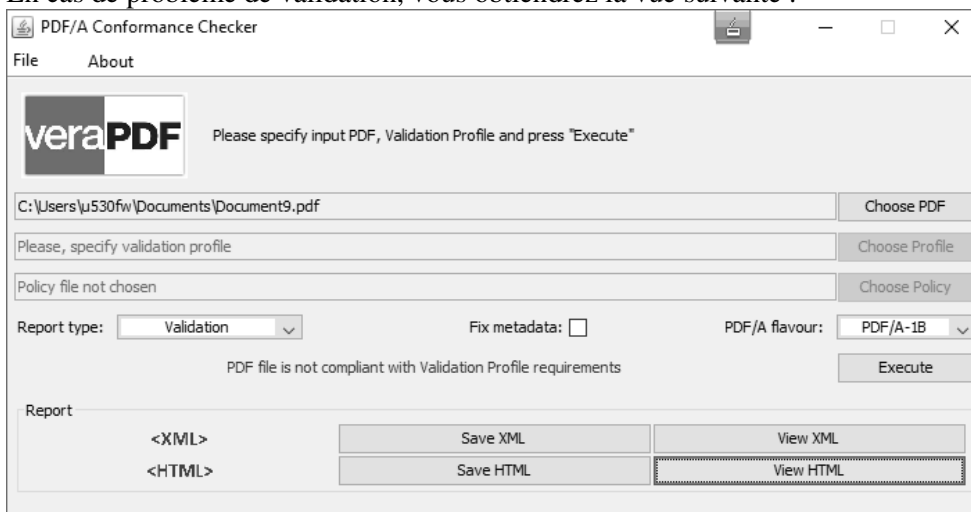
- Chargement du fichier PDF (en utilisant le bouton <Choose PDF>)
- Report type doit contenir <Validation>
- PDF/A flavour doit contenir <PDF/A-1B>

Lorsque ces éléments sont présents vous pouvez cliquer le bouton Execute.

Si le document est un PDF/A-1B valide vous obtiendrez le résultat suivant :



En cas de problème de validation, vous obtiendrez la vue suivante :



Dans ce cas de figure vous pouvez examiner le rapport HTML en cliquant sur <View HTML> pour visualiser les défauts détectés sur le PDF.



Validation Report

File: C:\Users\lu530fw\Documents\Document9.pdf
Validation Profile: PDF/A-1B validation profile
Compliance status: Failed

Statistics

Version: 1.2.2
Build Date: 2017-03-01T20:46:00+01:00
Processing time: 00:00:00:079
Total rules in Profile: 102
Passed Checks: 651
Failed Checks: 1

Explication de la clause qui n'a pas été validée
(Il est possible de cliquer sur ce lien pour obtenir une explication plus détaillée)

Validation information

Rule	Status
Specification: ISO 19005-1:2005, Clause: 6.7.11, Test number: 2	
The value of pdfaid:part shall be the part number of ISO 19005 to which the file conforms.	Failed
1 occurrences	Hide
PDFAIidentification	
part == 1	
root/document[0]/metadata[0](19 0 obj PDMetadata)/XMPPackage[0]/Identification[0]	

4.4. Création de documents PDF/A

Ce chapitre est à destination de développeurs d'applications souhaitant intégrer dans leur processus documentaire la création de PDF/A-1B valides. Il donne des exemple d'utilisation d'APIs permettant de fournir des PDF/A valides.

PDFLib (APIs commerciales)

L'exemple ci-dessous détaille des processus de création (langage JAVA ou PHP), mais ces APIs commerciales sont utilisables sur différents outils de développement et sur différentes plateformes :
<https://www.pdflib.com/pdflib-cookbook/pdfa/>

PDFBox (APIs Open-Source)

L'exemple suivant détaille un processus de création PDF/A-1B (langage JAVA) :
<https://pdfbox.apache.org/1.8/cookbook/pdfcreation.html>

iText (APIs en partie Open-Source)

L'exemple suivant détaille un processus de création PDF/A (langage JAVA) :
<http://developers.itextpdf.com/examples/archiving-and-accessibility-itext5/pdfa-1>