



Guide d'application

Table des matières

1	Préambule	2
2	Configurations système requises.....	3
3	Problèmes en suspens / Commentaires / Souhaits de modifications	3
4	Introduction double validation PDF/A	3
5	Données de diagnostic	4
5.1	PDF-Diagnosedaten.kost-val.xml	4
5.2	KaD-Diagnosedaten.kost-val.xml.....	4
6	Installation.....	5
6.1	Installation de KOST-Val.....	5
6.2	Installation optionnelle des applications tierces	5
7	Configuration de KOST-Val	7
7.1	Composants du fichier de configuration « kostval.conf.xml »	7
7.2	« KaD_SignatureFile_V70.01.xml » et « jhove.conf »	9
8	Ressources de KOST-Val.....	9
9	Exécuter la validation.....	10
9.1	Exécuter en utilisant le GUI « KOST-Val_fr.exe »	10
9.2	Démarrer la validation manuellement	12
10	Évaluer la validation.....	14
11	Description des étapes de validation	15
11.1	TIFF	15
11.2	SIARD	16
11.3	PDF/A	17
11.1	JP2.....	18
11.2	SIP	19
12	Droits d'auteur	20
12.1	« PDF/A Manager » de PDFTron.....	21
12.2	Licence « 3-Heights™ PDF/A Validator API »	21
13	Annexe.....	23
13.1	Propriétés du fichier de configuration « kostval.conf.xml »	23
13.2	Structure du programme	27

1 Préambule

KOST-Val est une application java pour la validation de la configuration et du contenu des fichiers TIFF (*Tagged Image File Format*), SIARD (*Software Independent Archiving of Relational Databases*), PDF/A, JP2 (JPEG 2000 ; *Joint Photographic Experts Group 2000*) ainsi que d'un paquet d'informations à verser (SIP ; *Submission Information Package*) pour le versement d'informations numériques. Cette application est sous licence GPL3+ et elle est mise à disposition du public par le CECO en Open Source. KOST-Val se base sur les composants inchangés d'autres fabricants qui sont intégrés directement dans le code source de KOST-Val. Les utilisateurs de KOST-Val sont tenus de suivre les dispositions de licence de tous ces composants. Des informations supplémentaires sont disponibles au chapitre 12.

KOST-Val remplit les exigences décrites ci-après.

Validation TIFF : KOST-Val lit un fichier TIFF et valide avec l'aide de Jhove la structure et le contenu ainsi que les caractéristiques clés telles que compression, espace colorimétrique et multipage. Les caractéristiques peuvent être configurées.

Validation SIARD : KOST-Val lit un fichier SIARD (eCH-0165 v1¹) et valide la structure et le contenu.

Validation PDF/A : KOST-Val lit un fichier PDF ou PDF/A (ISO 19005-1 et 19005-2) et valide avec l'aide du « 3-Heights™ PDF/A Validator » de PDF-Tools ou avec l'aide du « PDF/A-Manager » de PDFTron la structure et le contenu du fichier PDF. KOST-Val répartit les divers messages d'erreur dans les principaux groupes tels que polices, graphique et métadonnées. KOST-Val fournit seulement une version limitée² de « 3-Heights™ PDF/A Validator » de PDF-Tools.

Validation JP2 : KOST-Val lit un fichier JP2 (ISO 15444) et valide avec l'aide de Jpylyzer la structure et le contenu.

Validation SIP : KOST-Val lit un SIP (eCH-0160 v1³ ainsi que AFS v1⁴ et v4⁵) et valide les points suivants de la spécification SIP. Les différents points sont validés dans des groupes tels que la structure des données, la validation du schéma et validation de la somme de contrôle. Auparavant une validation du format est effectuée.

Les résultats (y compris les messages d'incohérences ou d'erreurs) de chaque étape sont écrits dans un fichier de log.

Les étapes de validation / les vérifications sont exécutées les unes après les autres. Lorsque cela est possible, la validation continue en cas d'erreur afin de réduire le nombre de cycles de correction.

Des informations plus détaillées sur les formats individuels et des étapes de validation sont présentées dans l'annexe.

¹ La spécification peut être téléchargée à partir du site eCH : <http://www.ech.ch/vech-web/page?p=dossier&documentNumber=eCH-0165>.

² La restriction est principalement liée à la capacité maximale de 72'000 pages par an. Plus d'informations sont sur la licence dans chapitre 12.2. Cette restriction peut être enlevée en acquérant une licence du « 3-Heights™ PDF/A Validator API » de PDF-Tools et en l'activant avec le License manager.

³ La spécification peut être téléchargée à partir du site eCH : <http://www.ech.ch/vech-web/page?p=dossier&documentNumber=eCH-0160>.

⁴ La spécification peut être téléchargée sur la page « L'interface de versement numérique » du site des AFS : <http://www.bar.admin.ch/themen/00876/00877/>.

⁵ La spécification peut être téléchargée sur la page « L'interface de versement numérique » du site des AFS : <http://www.bar.admin.ch/themen/00876/00877/01561/>.

2 Configurations système requises

Microsoft Windows 98 ou plus récente

Au moins 128 MB de RAM (512 MB ou plus recommandé)

Au moins 20 GB d'espace sur le disque dur

Java Runtime Environment (JRE), version 6 respectivement 1.6

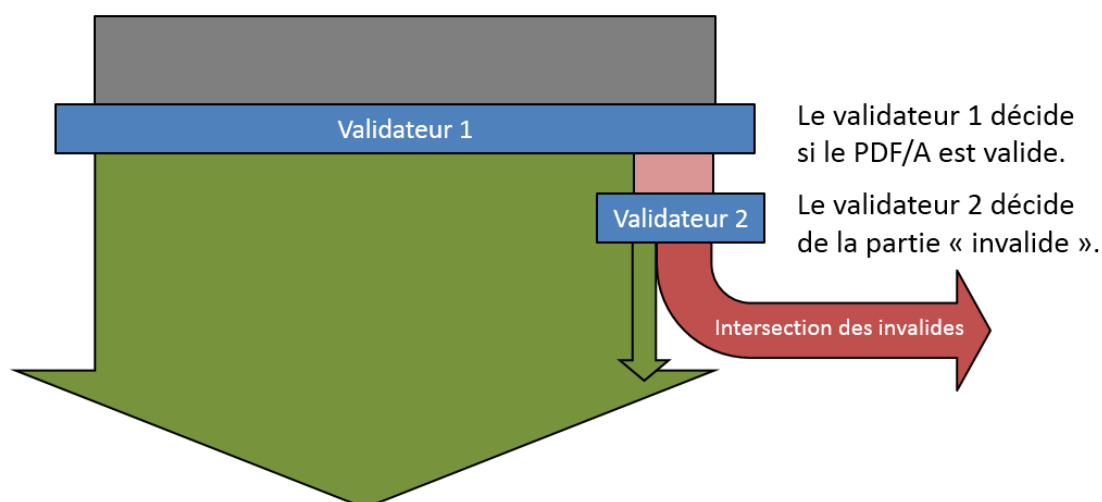
3 Problèmes en suspens / Commentaires / Souhaits de modifications

Les points ouverts (erreurs, bogues, souhaits de modifications) sont visibles sur la plate-forme de développement GitHub (<https://github.com/KOST-CECO/KOST-Val/issues>) et peuvent être rapportés à kost-val@kost-ceco.ch.

Cette liste peut être complétée par tous et est traitée par l'équipe de développement.

4 Introduction double validation PDF/A

Pour le PDF/A, le CECO propose la possibilité d'une double validation. Dans ce processus, un fichier PDF/A est d'abord contrôlé par une première validation. En cas de résultat invalide, un contrôle au moyen d'un deuxième validateur est effectué. Le PDF/A est considéré comme valide si au moins un des validateurs l'a identifié comme valide et comme invalide si les deux validateurs l'ont identifié comme invalide.⁶



La double validation PDF/A ne peut être utilisée que si les archives autorisent la prise en charge de fichiers PDF/A potentiellement invalides. Si ce n'est pas le cas, il ne faut pas faire la double validation.

⁶ C'est pourquoi la double validation ne peut être effectuée qu'avec des validateurs PDF/A de qualité. Les versions les plus récentes de « 3-Heights™ PDF/A Validator » de PDF Tools et de « PDF/A-Manager » de PDFTron répondent à ces exigences élevées.

Pour la double validation, il faut non seulement avoir une licence pour « 3-Heights™ PDF/A Validator » de PDF-Tools, mais également pour « PDF/A-Manager » de PDF-Tron, et il faut les avoir installés. Seule une version limitée de « 3-Heights™ PDF/A Validator » de PDF-Tools est fournie avec KOST-Val. Si seulement un validateur est disponible, il effectuera automatiquement une validation simple.

Le constat que même des validateurs PDF/A de qualité peuvent donner des résultats différents est à la base de la conception de la double validation. Cela tient d'une part au fait que la norme PDF/A originelle implique un ensemble d'autres normes et standards qui ne sont pas implémentés jusque dans les derniers détails de manière obligatoire. D'autre part, certains objectifs de la norme sont formulés de telle sorte qu'ils peuvent légitimement être implantés de différentes manières. Pouvoir implémenter de manière homogène et exhaustive tous les outils importants de la spécification est de la musique d'avenir. C'est pourquoi la double validation de KOST-Val offre une solution provisoire.

5 Données de diagnostic

KOST-Val crée dans certains cas différentes données de diagnostic. Ces dernières sont décrites ci-après et le chemin vers les données de diagnostic peut être au besoin adapté dans la configuration (voir aussi chapitre 7.1.1).

5.1 PDF-Diagnosedaten.kost-val.xml

Durant l'utilisation de la double validation PDF/A, en cas de désaccord des validateurs, KOST-Val recueille dans le fichier « PDF-Diagnosedaten.kost-val.xml » les données de diagnostic concernées, qui peuvent être utiles pour améliorer les outils et la norme. En plus du chemin de fichier complet, sont également recueillis la version PDF/A avec la conformité voulue et les messages d'erreur annonçant les désaccords.

Nous vous serions reconnaissants de bien vouloir envoyer à l'occasion au Bureau du CECO (kost-val@kost-ceco.ch) ces données de diagnostic PDF.

5.2 KaD-Diagnosedaten.kost-val.xml

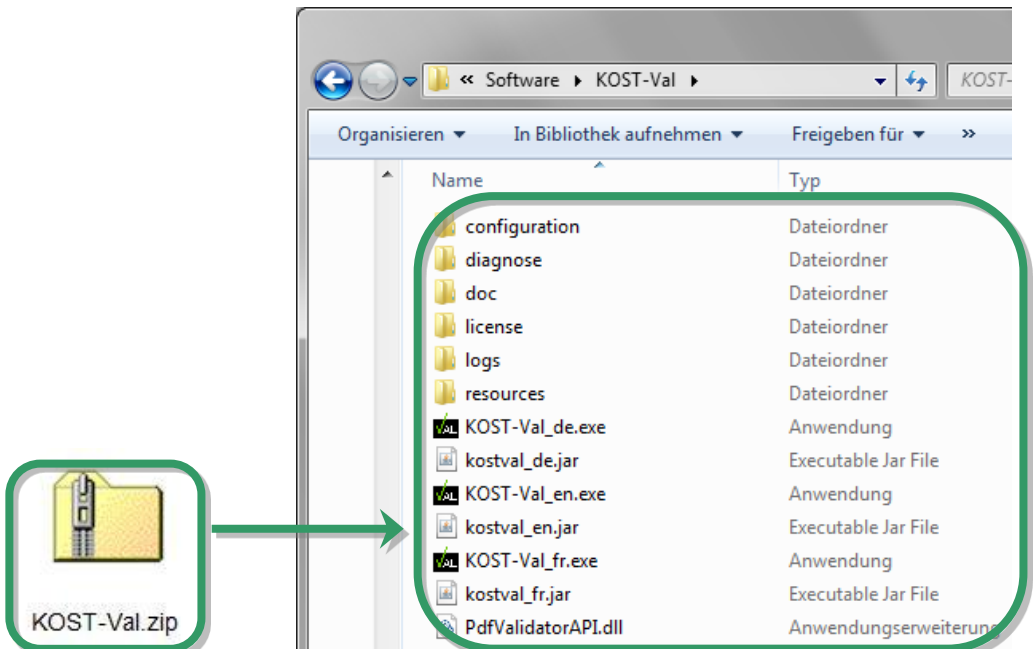
Durant l'utilisation de la validation SIP, KOST-Val recueille dans « KaD-Diagnosedaten.kost-val.xml » des informations sur les formats de fichier Cfa disponibles ainsi que sur leur nombre. Ceci permet d'apprendre quels formats de fichier sont contenus dans le dossier « content » des SIP validés et à quelle fréquence. Ce fichier peut être utilisé conformément au développement à la fois de KOST-Val (formats de fichier supplémentaires) et d'une planification de la préservation systématique avec une approche fondée sur les risques.

Nous vous serions reconnaissants de bien vouloir envoyer sur demande au Bureau du CECO (kost-val@kost-ceco.ch) ces fichiers de diagnostic Cfa.

6 Installation

6.1 Installation de KOST-Val

1 Télécharger KOST-Val et l'extraire dans le répertoire souhaité.



The screenshot shows a file explorer window with the address bar set to « Software » > KOST-Val. The file list is as follows:

Name	Typ
configuration	Dateiordner
diagnose	Dateiordner
doc	Dateiordner
license	Dateiordner
logs	Dateiordner
resources	Dateiordner
KOST-Val_de.exe	Anwendung
kostval_de.jar	Executable Jar File
KOST-Val_en.exe	Anwendung
kostval_en.jar	Executable Jar File
KOST-Val_fr.exe	Anwendung
kostval_fr.jar	Executable Jar File
PdfValidatorAPI.dll	Anwendungserweiterung

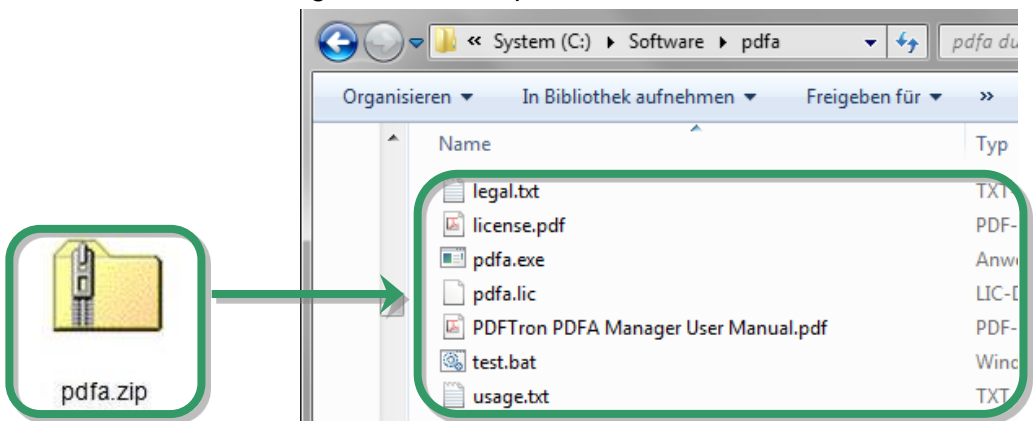
6.2 Installation optionnelle des applications tierces

6.2.1 Installation du « PDF/A Manager » de PDFTron

! Pour l'exécution d'une validation PDF/A, l'application indirecte du logiciel commercial « PDF/A Manager » de PDFTron a été mise en œuvre (au moyen de l'appel de la console) supplémentaire. Pour des tests, une version d'essai peut être téléchargée. Si « PDF/A-Manager » de PDFTron n'est pas installé, le « 3-Heights™ PDF/A Validator » intégrée de PDF-Tools est utilisé.

1 Télécharger la version en ligne de commande du « PDF/A Manager » (fichier ZIP) sur <http://www.pdftron.com/downloads.html>.

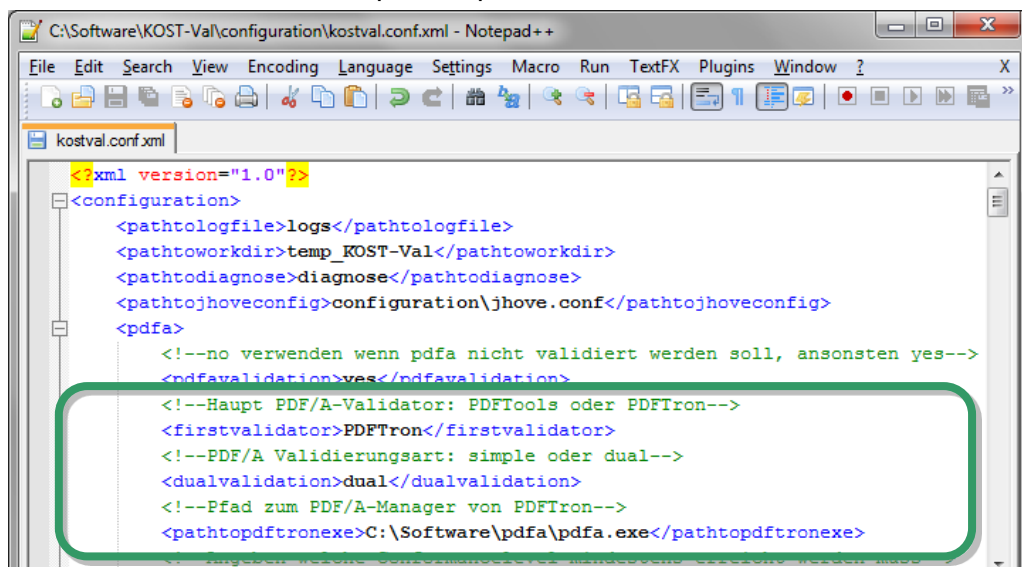
2 Extraire « PDF/A Manager » dans le répertoire souhaité.



The screenshot shows a file explorer window with the address bar set to « System (C:) » > Software > pdfa. The file list is as follows:

Name	Typ
legal.txt	TXT
license.pdf	PDF-
pdfa.exe	Anw
pdfa.lic	LIC-
PDFTron PDF/A Manager User Manual.pdf	PDF-
test.bat	Wind
usage.txt	TXT

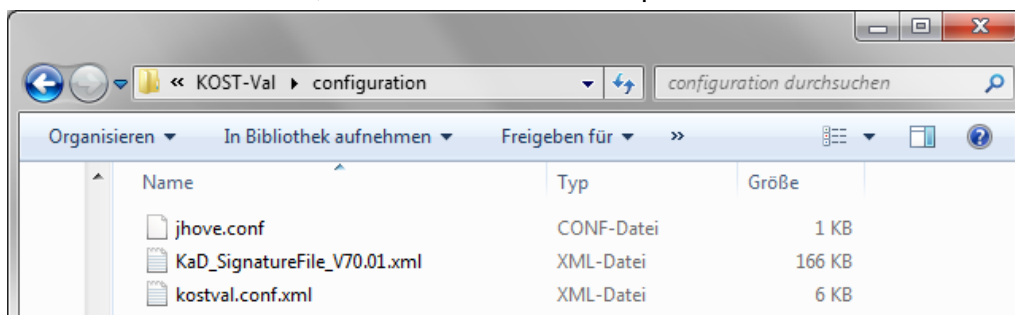
- 3 Compléter le fichier de configuration « KOST-Val\configuration\kostval.conf.xml » avec le lien vers pdfa.exe puis activer la double validation PDF/A en cas de besoin. En outre, le validateur principal « firstvalidator » doit être changé vers PDFTron à moins que la restriction d'utilisation de « 3-Heights™ PDF/A Validator » de PDF-Tools ait été enlevée. Utiliser un éditeur XML ou Notepad++⁷ pour ces modifications.



⁷ L'application portable Notepad++ peut être téléchargée à partir du site http://portableapps.com/de/apps/development/notepadpp_portable.

7 Configuration de KOST-Val

Toutes les configurations de KOST-Val peuvent être effectuées dans le fichier de configuration « kostval.conf.xml » dans le sous-dossier « configuration ». Pour des modifications d'un fichier XML, un éditeur XML ou Notepad++⁸ doit être utilisé.



Le fichier « KaD_SignatureFile_V70.01.xml » et le fichier « jhove.conf » sont aussi enregistrés dans le dossier « configuration ». Ils sont décrits dans le chapitre 7.2.

7.1 Composants du fichier de configuration « kostval.conf.xml »

Le fichier de configuration « kostval.conf.xml » est divisé en différentes parties, qui sont décrites en détail dans le chapitre 13.1.

La configuration délivrée permet la validation immédiate de PDF/A, TIFF, SIARD et SIP.

Ci-après une brève description des composants.

7.1.1 Partie générale

balise xml	Description : Valeur initiale
<pathtologfile>	Désignation du chemin vers le répertoire logs de KOST-Val : logs
<pathtoworkdir>	Chemin d'accès au répertoire de travail temporaire du KOST-Val : temp_KOST-Val
<pathtodiagnose>	Chemin vers le répertoire des données de diagnostic (voir aussi chapitre 5) : diagnose
<pathtojhoveconfig>	Désignation du chemin vers le fichier de configuration de Jhove : configuration\jhove.conf

7.1.2 Partie PDF/A

balise xml	Description : Valeur initiale
<pdfavalidation>	Indique si la validation PDF/A doit avoir lieu : yes
<firstvalidator>	Indique avec quel validateur la validation PDF/A principale doit être effectuée : PDFTools
<dualvalidation>	Indique si une double ou une simple validation doit être effectuée (voir aussi chapitre 4) : simple
<pathtopdftronexe>	Chemin d'accès à PDFTron (voir aussi chapitre 6.2.1) : C:\Software\pdfa\pdfa.exe

⁸ L'application portable Notepad++ peut être téléchargée à partir du site http://portableapps.com/de/apps/development/notepadpp_portable.

<pdfa1>	Indique quelle conformité doit être atteinte au moins dans la version 1 : 1B
<pdfa2>	Indique quelle conformité doit être atteinte au moins dans la version 2 : 2B

7.1.3 Partie SIARD

balise xml	Description : Valeur initiale
<siardvalidation>	Indique si la validation SIARD doit avoir lieu : yes

7.1.4 Partie JP2

balise xml	Description : Valeur initiale
<jp2validation>	Indique si la validation JP2 doit avoir lieu : yes

7.1.5 Partie TIFF⁹

balise xml	Description : Valeur initiale
<tiffvalidation>	Indique si la validation TIFF doit avoir lieu : yes
<allowedcompressionX>	Indique si l'algorithme de compression X est autorisé ou non. Les algorithmes suivants sont admises dans l'état initial : uncompressed, CCITT 1D, CCITT Group 3, CCITT Group 4, LZW, PackBits
<allowedphotointerY>	Indique si l'espace colorimétrique Y est autorisé ou non. Les espaces suivants sont admis dans l'état initial : white is zero, black is zero, RGB, palette color
<allowedbitspersampleZ>	Indique si Z bits per sample (par canal) est autorisé ou non. Les valeurs suivantes sont admises dans l'état initial : 1, 4, 8, 16
<allowedmultipage>	Indique si TIFF multipage est autorisé ou non : 1
<allowedtiles>	Indique si la structure en tuiles est autorisée ou non : 0
<allowedsize>	Indique si la taille du fichier de 1'000'000'000 bytes (~ 1 GB) et plus est autorisée ou non : 0

7.1.6 Partie SIP

balise xml	Description : Valeur initiale
<allowedlengthofpaths>	Nombre maximal de caractères autorisé dans la longueur du chemin : 179
<allowedlengthoffiles>	Nombre maximal de caractères autorisé dans le nom du fichier : 40
<allowedsipname>	Spécifications pour la composition du nom SIP : SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_\\w{3}

⁹ Les valeurs initiales sont fondées sur la recommandation Cfa - planification de la pérennisation (http://kost-ceco.ch/cms/index.php?preservation_fr), qui repose en partie sur le TIFF de base. Si nécessaire, les fichiers TIFF de base ont été précisés ou développés.

<pathtodroidssignature>	Chemin d'accès au fichier de signature DROID utilisé : configuration\KaD_SignatureFile_V70.01.xml
<allowedversionbar1>	Indiquant si la version 1 du AFS-SIP est autorisée : 1
<allowedversionbar4ech1>	Indiquant si la version 4 du AFS-SIP est autorisée respectivement eCH-0160 version 1 : 1
<allowedformats>	Récapitulatif des formats de fichiers autorisés, se composant de sous-éléments avec des informations détaillées. Les formats suivants sont admis dans l'état initial : TXT, TIFF, JP2, PDFa1, PDFa2, WAVE, CSV, SIARD

7.2 « KaD_SignatureFile_V70.01.xml » et « jhove.conf »

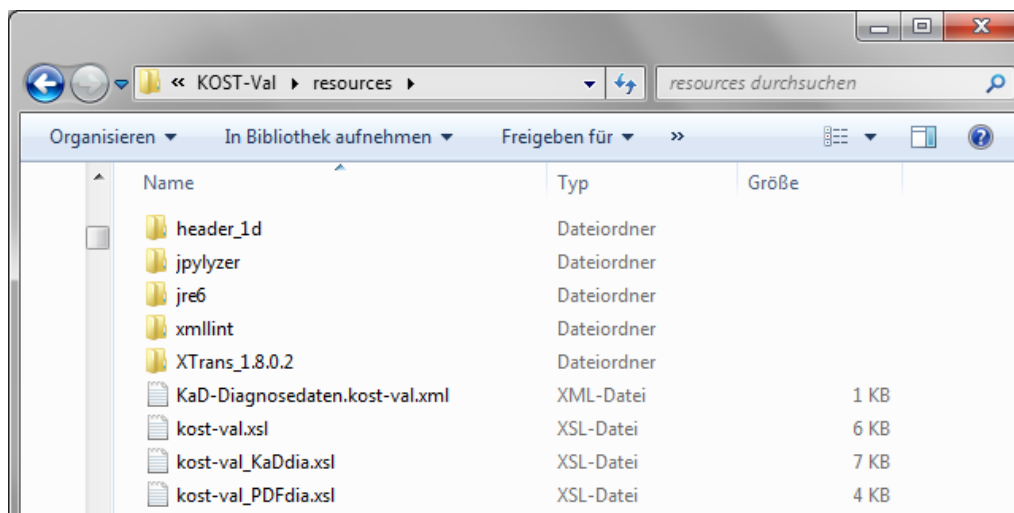
Le fichier « KaD_SignatureFile_V70.01.xml » et le fichier « jhove.conf » sont aussi enregistrés dans le dossier « configuration », qui n'a pas besoin d'être ajusté.

« KaD_SignatureFile_V70.01.xml » est nécessaire pour la reconnaissance du format. Il comprend l'état du DROID au 28.08.2012 et a été en plus adapté par le CECO dans le domaine de la reconnaissance des formats de données d'archivage¹⁰.

« jhove.conf » est requis pour la validation de Jhove intégré dans KOST-Val.

8 Ressources de KOST-Val

Toutes les ressources de KOST-Val sont stockées dans le sous-dossier « resources ».



¹⁰ A la place du fichier de signature de DROID, KOST-Val utilise un fichier de signature compatible développé par le CECO (voir aussi https://github.com/KOST-CECO/KaD_SignatureFile), KaD-SignatureFile. Ce dernier ne comprend que les formats analysés dans le Cfa et permet de les reconnaître avec la granularité recommandée par le CECO. Il aboutit ainsi à un gain significatif d'efficacité et à une facilité d'utilisation augmentée par rapport au fichier de signature de DROID.

9 Exécuter la validation

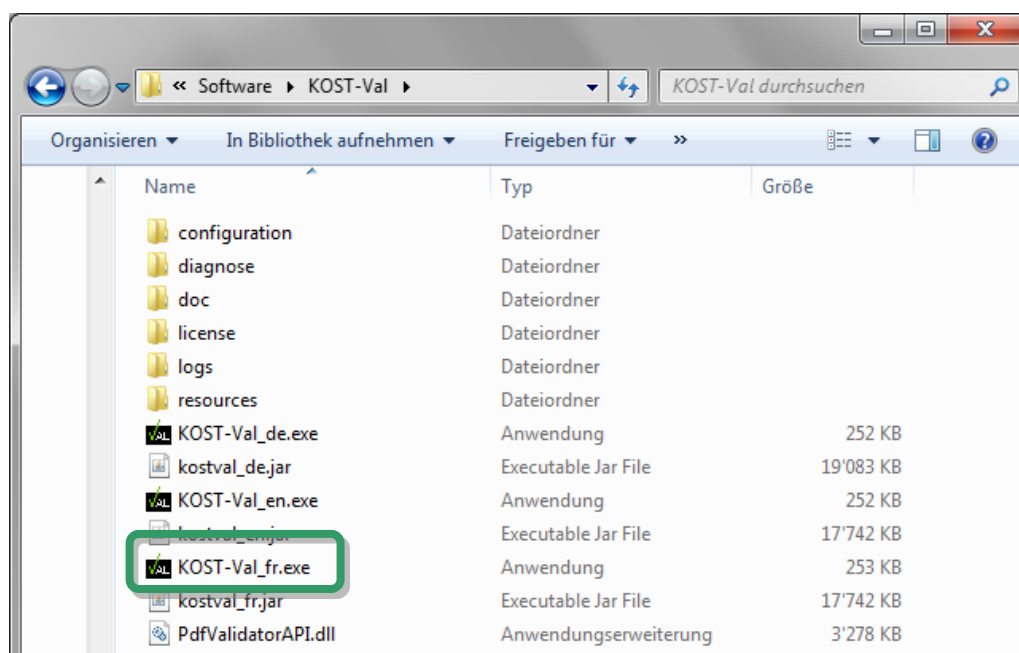


KOST-Val n'est pas thread-safe!

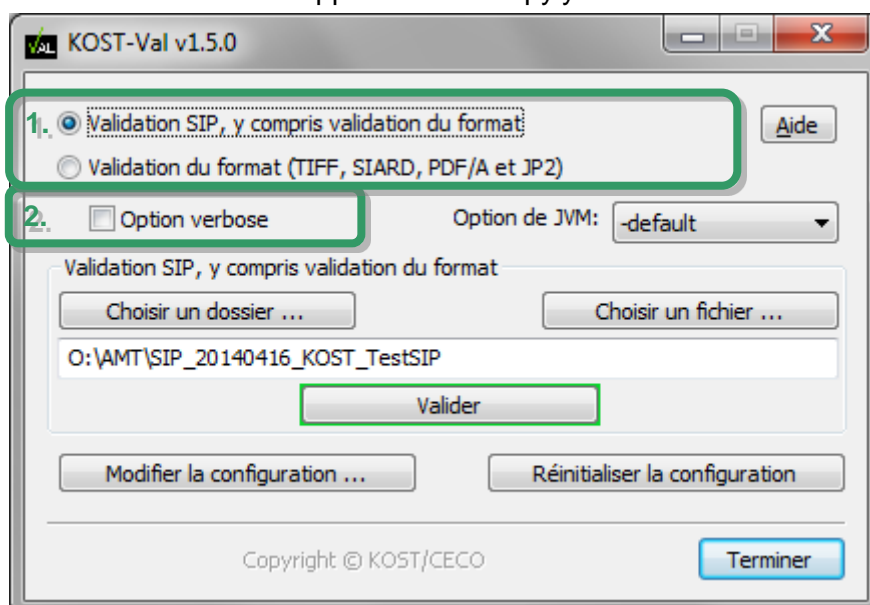
Cela signifie que KOST-Val ne peut pas être exécuté plusieurs fois simultanément sans interférences. L'exécution simultanée de plusieurs instances de KOST-Val peut entraîner des erreurs comme une copie de travail manquante.

9.1 Exécuter en utilisant le GUI « KOST-Val_fr.exe »

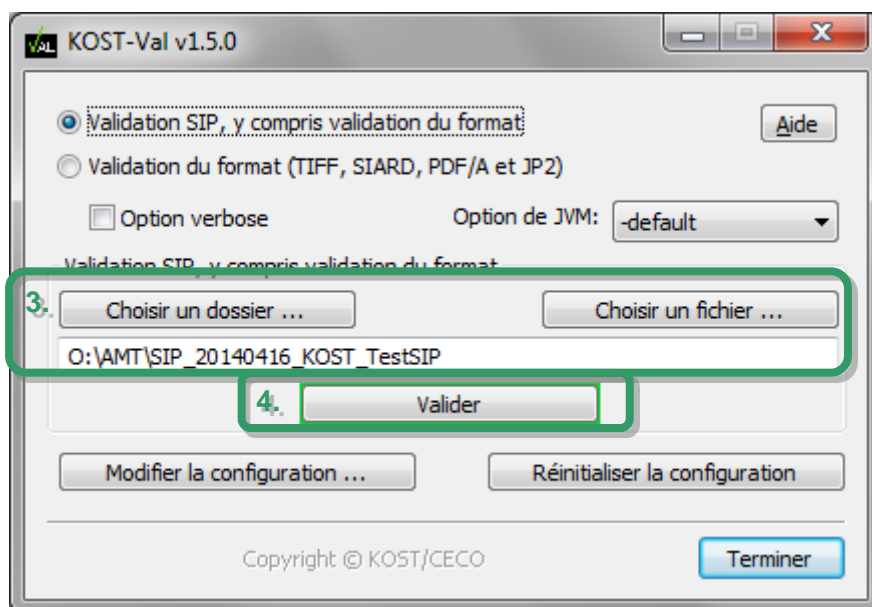
- 1 Lancer KOST-Val en double-cliquant sur « KOST-Val_fr.exe » dans le dossier « KOST-Val ».



- 2
 1. Choisir la validation de format ou la validation SIP (y compris la validation de format).
 2. Sélectionner si le rapport Jhove et Jpylyzer doit être conservé ou non.



3. Entrer directement le lien ou sélectionner le fichier ou le dossier à valider.
4. Lancer la validation.

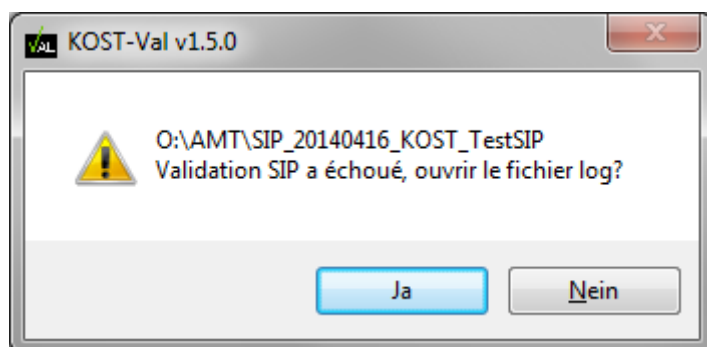


Note :

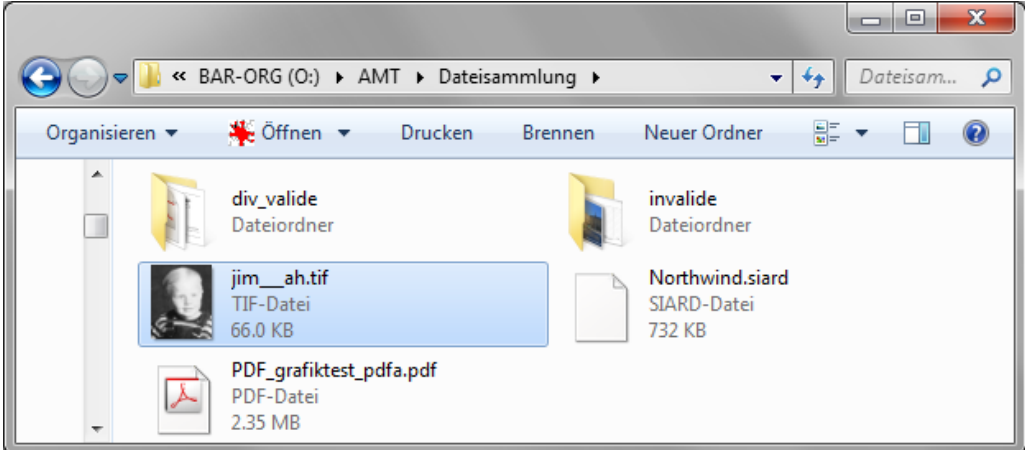
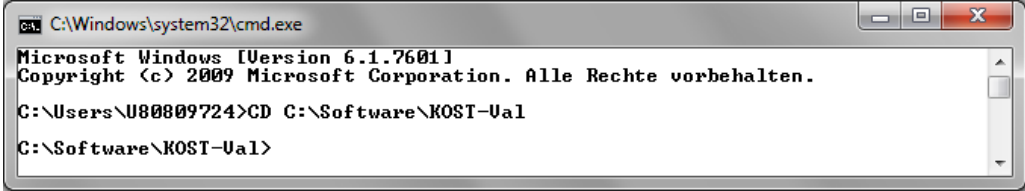
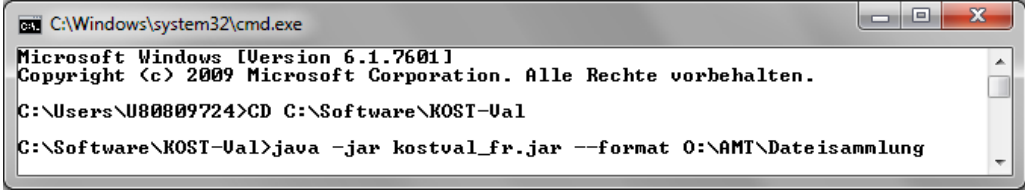
Si nécessaire, la valeur défaut de la mémoire virtuelle de Java peut être rapidement adaptée. -Xmx doit être adaptée sur « Out of Memory » et -Xss à des erreurs « Stack Overflow ».

Si on le souhaite, la configuration peut être adaptée provisoirement. La configuration provisoire est remise à son état initial en appuyant sur le bouton « Terminer » et « Réinitialiser la configuration »

- 3 A la fin de la validation, le résultat est affiché. Le fichier KOST-Val-Log peut alors être ouvert.



9.2 Démarrer la validation manuellement

1	<p>Pour valider mettre en place d'abord le répertoire ou le fichier.</p> 
2	<p>Ouvrir la fenêtre d'exécution (démarrer → Exécuter... → cmd [OK]) et aller dans le répertoire de travail (CD C:\Software\KOST-Val)¹¹.</p> 
3	<p>Appeler le programme KOST-Val (en séparant tous les composants d'entrée par des espaces).</p>  <p>A resources\jre6\bin\java.exe¹² -jar B kostval_fr.jar C --format (ou --sip pour la validation SIP) D O:\AMT\Dateisammlung ou O:\AMT\Dateisammlung\jim__ah.tif E « rien » ou -v (option verbose)</p> <p>A = appel de Java (resources\jre6\bin\java.exe -jar), B = lien relationnel et kostval_fr.jar, C = --format (ou --sip pour la validation SIP), D = lien relationnel et nom complet du répertoire ou fichier à valider, E = option verbose, si le rapport original doit être sauvegardé.</p> <p><u>Notes :</u> Quand un élément d'entrée contient des espaces, celui-ci doit être inscrit entre guillemets.</p>

¹¹ Le disque est par exemple changé avec CD /D.

¹² Utiliser la commande `java -jar` si le Java Runtime Environment (JRE) version 6 est la version standard. JRE7 peut causer des collisions. KOST-Val est massivement plus lent avec JRE8.

Il est également possible d'appeler KOST-Val depuis tout autre emplacement. Cependant, cela nécessite l'utilisation de liens absolus, non seulement dans l'appel du programme lui-même, mais aussi dans le fichier de configuration « kostval.conf.xml ».

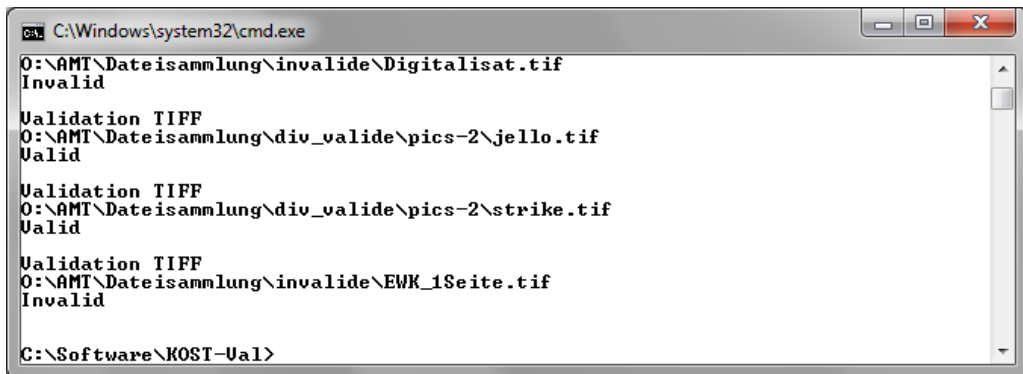


```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Alle Rechte vorbehalten.

C:\Users\U80809724>java -jar C:\Software\KOST-Val\kostval_fr.jar --format O:\AMT\Dateisammlung
```

- A** C:\Software\KOST-Val\resources\jre6\bin\java.exe -jar
- B** C:\Software\KOST-Val\kostval_fr.jar
- C** --format (ou --sip pour la validation SIP)
- D** O:\AMT\Dateisammlung ou O:\AMT\Dateisammlung\jim____ah.tif
- E** « rien » ou -v (option verbose)

- 4** La validation du fichier est terminée dès que « Valid » ou « Invalid » est affiché dans la fenêtre cmd. Le dossier a été validé une fois que le prompt (C:\Software\KOST-Val>) apparaît.



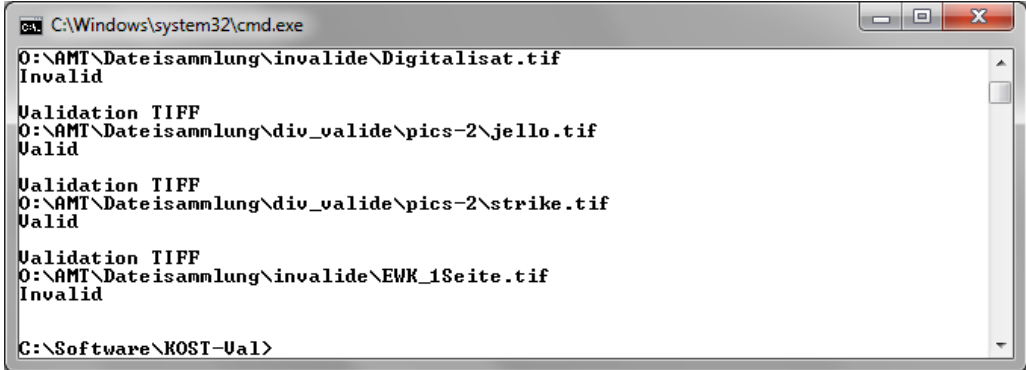
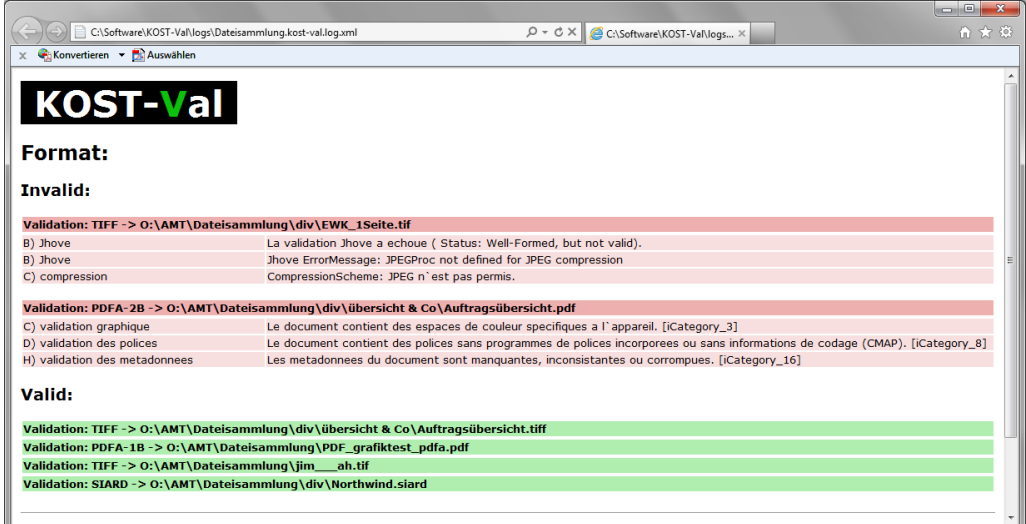
```
C:\Windows\system32\cmd.exe
O:\AMT\Dateisammlung\invalide\Digitalisat.tif
Invalid
Validation TIFF
O:\AMT\Dateisammlung\div_valide\pics-2\jello.tif
Valid
Validation TIFF
O:\AMT\Dateisammlung\div_valide\pics-2\strike.tif
Valid
Validation TIFF
O:\AMT\Dateisammlung\invalide\EWK_1Seite.tif
Invalid
C:\Software\KOST-Val>
```

Les résultats détaillés se trouvent dans le fichier « kost-val.log.xml ».

Le résultat de la validation globale (fichier correct/erroné) est aussi délivré et visible dans le statut *exit* du programme, afin que la validation puisse être liée dans la chaîne de traitement automatisée. Le statut *exit* peut prendre les valeurs suivantes :

- 0 tout est OK, valide
- 1 erreur dans l'appel du programme
- 2 échec de la validation, invalide

10 Évaluer la validation

1	Les étapes de validation sont décrites dans le chapitre 11.
2	<p>Après le démarrage de KOST-Val le type de validation et le fichier sont affichés. Si la validation a été réalisée sans erreurs, « Valid » est affiché.</p> <p>Si la validation a échoué, Invalid est affiché.</p>  <p>Si l'option « -v » (verbose pour « bavard ») est entrée, le rapport original est écrit dans le fichier log. L'option « -v » est uniquement intéressante pour les spécialistes dans de rares cas, puisque tous les messages d'erreur sont transférés à 100 % vers KOST-Val.</p>
3	<p>La sortie du log spécifie des informations détaillées supplémentaires sur les étapes de validation invalides, en particulier les étapes invalides avec l'erreur.</p> 

11 Description des étapes de validation

11.1 TIFF

Chaque étape de validation TIFF est décrite ci-après de façon détaillée.

A détection

Le fichier TIFF est-il reconnu comme tel ?

B Jhove

La validation Jhove est-elle réussie ?

C compression

La compression utilisée est-elle autorisée selon le fichier de configuration ?

D espace colorimétrique

L'espace colorimétrique utilisé est-il autorisé selon le fichier de configuration ?

E BitsPerSample

Les BitsPerSample utilisés sont-ils autorisés selon le fichier de configuration ?

F Multipage

L'utilisation de la propriété « multipage » est-elle autorisée selon le fichier de configuration ?

G tuiles

L'utilisation de la propriété « tuiles » est-elle autorisée selon le fichier de configuration ?

H grandeur

La taille du fichier est-elle autorisée selon le fichier de configuration ?

11.2 SIARD

Ce chapitre décrit plus en détail les étapes de validation. Ces critères sont un extrait des critères obligatoires de la spécification SIARD.

A lisibilité

Le fichier SIARD peut-il être lu ?

B validation de la structure des données

La structure primaire des données est-elle correcte ?

C validation du header

Le dossier header est-il valide ?

D validation de la structure

La structure de metadata.xml (folder) correspond-elle à la structure des fichiers de content ?

E validation des colonnes

Les données du fichier metadata.xml ont-elles été transférées correctement dans le fichier tableZ.xsd?

F validation des lignes

Les données du fichier metadata.xml ont-elles été transférées correctement dans le fichier tableZ.xsd?

G validation des tables

Les noms de colonnes sont-ils uniques dans la table ?

H validation du contenu

Les fichiers XML dans le contenu sont-ils valides par rapport à leur définition de schéma (XSD) ?

I détection du format SIARD

Le fichier SIARD est-il détecté comme SIARD ?

J données primaires supplémentaires

Tous les fichiers ou dossiers contenu dans le dossier « content » sont-ils décrits dans metadata.xml?

11.3 PDF/A

Les messages du rapport initial de PDFTron sont attribués aux modules basés sur le premier chiffre de « Error ID ». Les messages d'erreur (« Message ») sont pris en charge 1:1, mais ils sont traduits (original en anglais). Lors de la validation avec PDF-Tools, les catégories sont attribuées aux modules, et le texte est affiché.

A validation générale

La validation des conditions générales telles que fichier PDF/A non corrompu est-elle réussie ?

B validation de la structure

La validation de la structure (par exemple BOF et EOF) est-elle réussie ?

C validation graphique

La validation graphique (par exemple, espace de couleur valide) est-elle réussie ?

D validation des polices

Les polices sont-elles intégrées correctement dans le fichier PDF/A (par exemple, les caractères utilisés sont-ils tous intégrés) ?

E validation de la transparence

Les exigences en matière de transparence sont-elles remplies pour le fichier PDF/A ?

F validation des annotations

Le fichier PDF/A ne contient-il que des annotations licites (p.ex. pas d'annotation 3D) ?

G validation des actions & interactions

Le fichier PDF/A ne contient-il que des actions licites (p.ex. pas de JavaScript) ?

Le fichier PDF/A ne contient-il que des interactions licites ?

H validation des métadonnées

Les métadonnées sont-elles correctement incluses dans le fichier PDF/A ?

I validation de la disponibilité

Les exigences à l'accessibilité (niveau de conformité A) sont-elles remplies ?

11.1 JP2

Les tests échoués de Jpylyzer sont attribués aux modules. Des messages d'erreur correspondant à ces derniers sont émis. Ce chapitre décrit plus en détail les étapes de validation de JP2.

A Détection et Jpylyzer

Le fichier JP2 est-il reconnu comme tel ? La validation avec Jpylyzer a-t-elle été réussie ?

B Validation des métadonnées

Les métadonnées ont-elles été saisies correctement ?

C Validation de l'image

L'image est-elle construite correctement ?

D Validation étendue

Les tests Jpylyzer additionnels ont-ils été réussis ?

11.2 SIP

Ce chapitre décrit plus en détail les étapes de validation. Ces critères sont un extrait des critères obligatoires de la spécification SIP.

Module 1 : Consistance paquet et XML

1A lisibilité

Le SIP (ZIP / ZIP64) peut-il être ouvert sans erreur ?

1B validation de la structure des données

Existe-t-il une structure de répertoire primaire correcte ?

1C nommage de la structure et des fichiers

Les noms correspondent-ils aux restrictions de la spécification ?

1D validation du schéma metadata.xml

Les métadonnées correspondent-elles aux fichiers schéma dans header/xsd ?

1E déterminer le type du SIP

Le type SIP est transmis et affiché : GEVER ou FILE.

1F existence des données primaires

Existe-t-il des fichiers primaires dans le répertoire content/ ?¹³

Module 2 : Consistance du fichier

2A données primaires manquantes

Tous les fichiers référencés existent-ils ?

2B données primaires supplémentaires

N'existe-t-il pas de fichiers primaires dans le SIP ?

2C validation de la somme de contrôle

Les sommes de contrôle des fichiers sont-elles correctes ?

2D description des données primaires GEVER

Tous les fichiers référencés sont-ils aussi enregistrés dans ordnungssystem ?

Module 3 : Consistance du format de fichier et de la date

3A détection des formats

Effectuer une reconnaissance et comparer si les formats sont autorisés.

3B formats supplémentaires

Lister tous les fichiers avec un format non spécifié et indiquer le format.

3C validation de formats

Les formats sont-ils valides ?

3D validation de période

Les indications temporelles dans (metadata.xml)/ablieferung sont-elles consistantes ?

¹³ Les fichiers primaires peuvent manquer dans un GEVER-SIP qui est utilisé uniquement pour l'archivage d'une structure de système de classement, mais jamais dans un FILE-SIP.

12 Droits d'auteur

KOST-Val a été développé par le CECO. Le CECO en possède tous les droits. KOST-Val a été publié par le CECO en 2012 sous la licence GNU General Public License v3+.

Notice :	This product includes software developed by the Apache Software Foundation (http://www.apache.org/).
-----------------	---

KOST-Val se base sur les composants inchangés suivants d'autres fabricants qui sont intégrés directement dans le code source de KOST-Val :

Programme tiers/Composants tiers	Version	Licence
3-Heights™ PDF/A Validator API http://www.pdf-tools.com		Voir chapitre 12.2
Apache Commons http://commons.apache.org/ - commons-collections-3.2.1.jar - commons-configuration-1.6.jar - commons-digester-1.8.jar - commons-lang-2.4.jar - commons-logging-1.1.1.jar	3.2.1 1.6 1.8 2.4 1.1.1	Apache License 2.0
Apache log4j http://logging.apache.org/log4j/	1.2.12	Apache License 2.0
Apache log4j http://logging.apache.org/log4j/	1.2.12	Apache License 2.0
Apache Xalan-Java http://xml.apache.org/xalan-j/	2.7.0	Apache License 2.0
Apache Xerces http://xerces.apache.org/	2.7.1	Apache License 2.0
DROID http://sourceforge.net/apps/mediawiki/droid/	5.0.3	3c BSD License
Jdom 2.0.0 http://www.jdom.org/	2.0.0	jdom License
Jhove http://hul.harvard.edu/jhove/	1.5	LGPL v2.1 License
Junit 4.4 http://www.junit.org/	4.4	CPL v1.0
Spring Framework API http://static.springsource.org/spring/docs/3.0.x/api/	3.0.0	Apache License 2.0
zip64 http://sourceforge.net/projects/zip64file/	1.02	GPL v2+ License

KOST-Val GUI utilise les composants suivants d'autres fabricants :

Programme tiers/Composants tiers	Version	Licence
Jpylyzer http://jpylyzer.openpreservation.org/	1.13.0	LGPL v3.0 License
NSIS v2.46 http://nsis.sourceforge.net/Main_Page	2.46	zlib/libpng License
XML.nsh http://nsis.sourceforge.net/XML_plug-in	2.0	zlib/libpng License
XmlLint http://xmlsoft.org/xmlLint.html	20630	MIT License
XTrans http://sourceforge.net/projects/xtrans/	1.8.0.2	GPL v2 License

Les utilisateurs de KOST-Val sont tenus de suivre les dispositions de licence de tous ces composants qui se trouvent dans le répertoire KOST-Val\license.

12.1 « PDF/A Manager » de PDFTron

Pour l'exécution d'une validation PDF/A, l'application indirecte de la version test du logiciel commercial PDF/A Manager de PDFTron a été mise en œuvre (au moyen de l'appel de la console). Ce logiciel n'est pas un composant de KOST-Val et doit être fourni séparément et installé pour une éventuelle utilisation. Des informations supplémentaires sont disponibles au chapitre 6.2.1.

12.2 Licence « 3-Heights™ PDF/A Validator API »

Pour l'utilisation de la version limitée de « 3-Heights™ PDF/A Validator » de PDF-Tools, le CECO a convenu des conditions de licence suivantes avec PDF Tools :

Anhang

zu den Allgemeinen Lizenzbedingungen der PDF Tools AG vom 01.03.2012,
Version 2.3

zwischen

PDF Tools AG
Geerenstrasse 33
CH-8185 Winkel

und

KOST
c/o Schweizerisches Bundesarchiv
Archivstrasse 24
CH-3003 Bern

(nachfolgend "**PDF TOOLS**" genannt)

(nachfolgend "**KOST**" genannt)

1. Vorbemerkung

Die untenstehenden Bedingungen sind Ergänzungen der Allgemeinen Lizenzbedingungen der PDF TOOLS und werden mit Unterzeichnung dieser Vereinbarung integrierter Vertragsbestandteil.

2. Individuelle Vereinbarung

Dieses Vertragsverhältnis regelt die Client-Lizenz zwischen der PDF TOOLS als Lizenzgeber und der KOST als Lizenznehmer gemäss nachfolgenden Spezialbestimmungen:


- PDF Tools AG erteilt für KOST eine kostenfreie OEM-Lizenz für das 3-Heights™ PDF/A Validator API als Zusatzfunktion ihrer eigenen Validator-Software (KOST-Val).
- Die Lizenz schliesst den Gebrauch der Software (KOST-Val) durch Gedächtnisinstitutionen, bestehend aus Archiven oder Bibliotheken, deren Zulieferer und der KOST selbst, ein.
- Der OEM-Lizenzschlüssel, welcher fest in KOST-Val eingebunden ist, darf nicht ausserhalb der Applikation (KOST-Val) verwendet werden.
- Die Lizenz ist zeitlich unbegrenzt, jedoch bezüglich Durchsatz pro Installation begrenzt (72'000 Seiten pro Jahr).
- Für die Verteilung der Software (KOST-Val) an den Anwender ist die KOST zuständig.
- Der First Level Support der Anwender erfolgt durch KOST. Second Level Support Fälle leitet KOST an PDF Tools AG weiter.
- Wenn der Anwender weitergehende Bedürfnisse hat, z.B. höherer Durchsatz, Integration in andere Applikationen etc. kauft er die Software (3-Heights™ PDF/A Validator API) direkt bei PDF Tools AG.
- Die KOST darf weiterhin den Quellcode von KOST-Val Open Source publizieren und KOST-Val gratis und ohne Registrierung abgeben.

3. Kundenreferenz

PDF TOOLS ist berechtigt, KOST als Kundenreferenz aufzuführen und die Zusammenarbeit und die Anwendung des Produktes der PDF TOOLS bei KOST zu beschreiben und zu publizieren.

Winkel, 7.10.2014

PDF Tools AG


Dr. Hans Bärnuss

Bern, 29. Sept. 2014

KOST


Für den Steuerungsausschuss
Dr. Beat Gnädinger

Les points suivants sont déterminants pour l'utilisateur :

- La licence comprend l'utilisation du logiciel (KOST-Val) par des institutions patrimoniales composées des archives ou des bibliothèques, leurs fournisseurs et le CECO lui-même.
- La clé de licence OEM qui est fermement liée à KOST-Val ne peut pas être utilisée hors de l'application (KOST-Val).
- La licence est illimitée dans le temps, mais le débit par installation est limité (72'000 pages par an).
- Le support de niveau 1 aux usagers est effectué par le CECO. Le CECO transmet à PDF Tools AG les cas relevant du support de niveau 2.
- Si l'utilisateur a d'autres besoins, par exemple un débit plus élevé, intégration dans d'autres applications, etc., il doit acheter le logiciel « 3-Heights™ PDF/A Validator API » directement auprès de PDF Tools AG.

Les utilisateurs de KOST-Val sont tenus de respecter les présentes dispositions de la licence.

13 Annexe

13.1 Propriétés du fichier de configuration « kostval.conf.xml »

Le fichier de configuration « kostval.conf.xml » est divisé en différentes parties, qui sont décrites en détail dans ce chapitre.

13.1.1 Partie générale

<pathlogfile>

Chemin d'accès au répertoire de log du KOST-Val. La valeur initiale est **logs**. Il est possible de saisir le chemin absolu ou le chemin relationnel comme dans la valeur initiale (du point de vue de kostval_fr.jar). Ce dossier ne doit pas être créé.

<pathtoworkdir>

Chemin d'accès au répertoire de travail temporaire du KOST-Val. La valeur initiale est **temp_KOST-Val**. Il est possible de saisir le chemin absolu ou le chemin relationnel comme dans la valeur initiale (du point de vue de kostval_fr.jar). Ce dossier ne doit pas être créé.

<pathtodinose>

Chemin vers le répertoire des données de diagnostic (voir aussi chapitre 5). La valeur initiale est **dinose**. Il est possible de saisir le chemin absolu ou le chemin relationnel comme dans la valeur initiale (du point de vue de kostval_fr.jar). Ce dossier ne doit pas être créé.

<pathtojhoconfig>

Chemin d'accès au fichier de configuration de JHOVE. La valeur initiale est **configuration\jho.conf**. Il est possible de saisir le chemin absolu ou le chemin relatif comme dans la valeur initiale (du point de vue de kostval_fr.jar).

13.1.2 Partie PDF/A

<pdfavalidation>

Définition si la validation PDF/A doit avoir lieu (yes) ou non (no). La valeur initiale est **yes**.

<firstvalidator>

Indique avec quel validateur la validation PDF/A principale doit être effectuée. Les validateurs PDF/A suivants sont mis en œuvre : 3-Heights™ PDF/A Validator de PDF-Tools (PDFTools) et PDF/A Manager de PDFTron (PDFTron). La valeur initiale est **PDFTools**.

<dualvalidation>

Indique si une validation double (dual) ou simple (simple) doit être effectuée (voir aussi chapitre 4). La valeur initiale est **simple**.

<pathtopdftronexe>

Chemin d'accès à PDFTron. La valeur initiale est **C:\Software\pdfa\pdfa.exe**. Il est possible de saisir le chemin absolu ou le chemin relatif comme dans la valeur initiale (du point de vue de kostval_fr.jar).

<pdfa1>

Indique quelle conformité doit être atteinte au moins dans la version 1. Il peut être choisi entre *Accessible* (1A), *Basic* (1B) ou non permis (no). La valeur initiale est **1B**.

<pdfa2>

Indique quelle conformité doit être atteinte au moins dans la version 2. Il peut être choisi entre *Accessible* (2A), *Basic* (2B), *Unicode* (2U) ou non permis (no). La valeur initiale est **2B**.

13.1.3 Partie SIARD

<siardvalidation>

Définition si la validation SIARD doit avoir lieu (yes) ou non (no). La valeur initiale est **yes**.

13.1.4 Partie JP2

<jp2validation>

Définition si la validation JP2 doit avoir lieu (yes) ou non (no). La valeur initiale est **yes**.

13.1.5 Partie TIFF

Les valeurs initiales sont fondées sur la recommandation Cfa - planification de la pérennisation (http://kost-ceco.ch/cms/index.php?preservation_fr), qui repose en partie sur le TIFF de base. Si nécessaire, les fichiers TIFF de base ont été précisés ou développés.

<tiffvalidation>

Définition si la validation TIFF doit avoir lieu (yes) ou non (no). La valeur initiale est **yes**.

<allowedcompressionX>

Indique si l'algorithme de compression est autorisé (description) ou non (0). Les valeurs initiales sont **uncompressed, CCITT 1D, CCITT Group 3, CCITT Group 4, LZW, PackBits**.

X	Description	Valeur initiale	Origine
1	uncompressed	uncompressed	Baseline-TIFF
2	CCITT 1D	CCITT 1D	Baseline-TIFF
3	CCITT Group 3	CCITT Group 3	recommandation Cfa-PP
4	CCITT Group 4	CCITT Group 4	recommandation Cfa-PP
5	LZW	LZW	recommandation Cfa-PP
7	ISO JPEG	0	Baseline-TIFF
8	Deflate	0	Baseline-TIFF
32773	PackBits	PackBits	Baseline-TIFF

<allowedphotointerY>

Indique si l'espace colorimétrique est autorisé (description) ou non (0). Les valeurs initiales sont **white is zero, black is zero, RGB, palette color**.

Y	Description	Valeur initiale	Origine
0	white is zero	white is zero	Baseline-TIFF
1	black is zero	black is zero	Baseline-TIFF
2	RGB	RGB	Baseline-TIFF
3	palette color	palette color	Baseline-TIFF

4	transparency mask	0	Baseline-TIFF
5	CMYK	0	Baseline-TIFF
6	YCbCr	0	Baseline-TIFF
7	CIE L*a*b*	0	Baseline-TIFF

<allowedbitspersampleZ>

Indique si le *bits per sample* (par canal) est autorisé (Z) ou non (0). Les valeurs initiales sont **1, 4, 8, 16**.

Z	Description	Valeur initiale	Origine
1	1 bits per sample (par canal)	1	recommandation Cfa-PP
2	2 bits per sample (par canal)	0	Baseline-TIFF
4	4 bits per sample (par canal)	4	Baseline-TIFF
8	8 bits per sample (par canal)	8	Baseline-TIFF
16	16 bits per sample (par canal)	16	recommandation Cfa-PP
32	32 bits per sample (par canal)	0	Baseline-TIFF
64	64 bits per sample (par canal)	0	Baseline-TIFF

<allowedmultipage>

Indique si TIFF multipage est autorisé (1) ou non (0). La valeur initiale est **1** et correspond à la recommandation de la planification de la pérennisation Cfa TIFF.

<allowedtiles>

Indique si la structure en tuiles est autorisée (1) ou non (0). La valeur initiale est **0** et répond aux exigences des Baseline TIFFs.

<allowedsize>

Indique si une taille du fichier de 1'000'000'000 bytes (~ 1 GB) et plus est autorisée (1) ou non (0). La valeur initiale est **0** et correspond à la recommandation de la planification de la pérennisation Cfa TIFF.

13.1.6 Partie SIP

<allowedlengthofpaths>

Nombre maximal de caractères autorisé dans la longueur du chemin. La valeur initiale est **179** et correspond à la restriction de la spécification du SIP : moins de 180 caractères.

<allowedlengthoffiles>

Nombre maximal de caractères autorisé dans le nom du fichier. La valeur initiale est **40** et correspond à la restriction du SIP des AFS (version 1) : moins de 40 caractères. En validant les nouvelles versions du SIP, il n'existe aucune restriction, donc ce champ ne concerne que la version 1 du SIP des AFS.

<allowsipname>

Spécifications pour la composition du nom du SIP. La valeur initiale (expression Regex) est **SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_\\w{3}** et correspond à la restriction de la spécification du SIP : SIP_[dateversement]_[Nom du service versant](_[Référence]), pour lequel la date doit être saisie au format AAAAMMJJ et la référence peut être utilisée de manière optionnelle.

<pathtodroidsignature>

Chemin d'accès au fichier de signature conforme à DROID utilisé dans le répertoire. La valeur initiale est **configuration\KaD_SignatureFile_V70.01.xml**. Des informations supplémentaires sont disponibles au chapitre 7.2.

<allowedversionbar1>

Indiquant si la version 1 du AFS-SIP est permise (1) ou non (0). La valeur initiale est **1**.

<allowedversionbar4ech1>

Indiquant si la version 4 du AFS-SIP respectivement eCH-0160 version 1 est permis (1) ou non (0). La valeur initiale est **1**.

<allowedformats>

Récapitulatif des formats de fichiers autorisés, se composant de sous-balises avec des informations détaillées. Les valeurs initiales sont visibles dans le tableau suivant. La liste détaillée est conservée intentionnellement, de sorte que les formats non autorisés puissent être facilement supprimés et que le moins de formats possible aient à être ajoutés. Les indications concernant l'extension et le PUID sont visibles dans KaD_SignatureFile_V70.01.xml.

<puid> </puid>	<extension> </extension>	<comment> </comment>
TXT	txt	Plain Text file
TIFF	tif, tiff	Tagged Image File Format
JP2	jp2, jpg2, j2c	JPEG2000
PDFA1	pdf	PDF/A-1
PDFA2	pdf	PDF/A-2
WAVE	wav, wave	Waveform Audio
CSV	csv	Comma Separated Values
SIARD	siard	SIARD
XML	xml	Ext. Markup Language
XSD	xsd	XML Schema Definition
JPEG	jpeg, jpg, jpe	JPEG
MP3	mp3	MPEG 1/2 Audio Layer 3
MP4	mp4, m4v, m4a, f4v, f4a	MPEG-4 Media File

13.2 Structure du programme

KOST-Val a été développé selon les exigences suivantes :

Exigences fonctionnelles :

Validation TIFF : KOST-Val lit un fichier TIFF et valide avec l'aide de Jhove les points suivants :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	détection
B	Jhove
C	compression
D	espace colorimétrique
E	BitsPerSample
F	multipage
G	tuiles
H	grandeur

Validation SIARD : KOST-Val lit un fichier SIARD et valide les points suivants :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	lisibilité
B	validation de la structure des données
C	validation du header
D	validation de la structure
E	validation des colonnes
F	validation des lignes
G	validation des tableaux
H	validation du contenu
I	détection du format SIARD
J	données primaires supplémentaires

Validation PDF/A : KOST-Val lit un fichier PDF et valide avec l'aide du 3-Heights™ PDF/A Validator de PDF-Tools ou avec l'aide du PDF/A-Manager de PDFTron les points suivants :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	validation générale
B	validation de la structure
C	validation graphique
D	validation des polices
E	validation de la transparence
F	validation des annotations & interactions
G	validation des actions
H	validation des métadonnées
I	validation de la disponibilité

Validation JP2 : KOST-Val lit un fichier JP2 et valide avec l'aide de Jpylyzer les points suivants :

Étape de validation	Nom de l'étape
A (abandon possible en cas d'erreur)	Détection et Jpylyzer
B	validation des métadonnées
C	validation de l'image
D	validation étendue

Validation SIP : KOST-Val lit un SIP et valide les points suivants :

Étape de validation	Nom de l'étape
1a (abandon en cas d'erreur)	lisibilité
1b (abandon en cas d'erreur)	validation de la structure des données
1c (abandon en cas d'erreur)	nommage de la structure et des fichiers
1d (abandon en cas d'erreur)	validation du schéma metadata.xml
1e	déterminer le type du SIP
1f	existence des données primaires
2a	données primaires manquantes
2b	données primaires supplémentaires
2c	validation de la somme de contrôle
2d	description des données primaires GEVER
3a	détection des formats
3b	formats supplémentaires
3c	validation de format
3d	validation de période

Les résultats (y compris les messages d'incohérences ou d'erreurs) de chaque étape sont écrits dans un fichier de log.

Le résultat de la validation globale (fichier correct/erroné) est aussi délivré et visible dans le statut *exit* du programme, afin que la validation puisse être liée dans la chaîne de traitement automatisée. Le statut *exit* peut prendre les valeurs suivantes :

- 0 tout est OK, valide
- 1 erreur dans l'appel du programme
- 2 échec de la validation, invalide

Les étapes de validation / les vérifications sont exécutées les unes après les autres. Lorsque cela est possible, la validation continue en cas d'erreur afin de réduire le nombre de cycles de correction.

Exigences non fonctionnelles :

Pour des tâches particulières, des programmes externes ou des frameworks Java appropriés sont utilisés.

L'application est constituée de manière modulaire, de sorte que la mise en place d'un ou plusieurs module(s) de validation supplémentaire(s) nécessite moins d'effort.

La sortie du programme et le fichier log permettent une lecture aisée du résultat de chaque validation et par conséquent l'utilisation de l'outil dans une chaîne de processus.

La sortie de la console se limite à la description des étapes de validation, les résultats valides ou invalides, le résultat global, ainsi que les chemins vers le fichier et le fichier log généré. Toutes les informations additionnelles sont spécifiées dans le fichier log.