Une entreprise commune des Archives suisses



Table des matières

1	Préa	mbule				
2		igurations système requises				
3	Problèmes en suspens / Commentaires / Souhaits de modifications					
4		duction à la double validation PDF/A				
5	Insta	ıllation de KOST-Val	4			
6	Configuration de KOST-Val					
	6.1	Composants du fichier de configuration « kostval.conf.xml »	4			
7	Ress	sources de KOST-Val	8			
8	Exéc	cuter la validation	<u>C</u>			
	8.1	KOST-Val avec l'interface graphique (GUI)	<u>C</u>			
	8.2	Démarrer la validation manuellement	11			
9	Droits d'auteur					
	9.1	Licence « 3-Heights™ PDF/A Validator API »	13			
		Licence « pdfaPilot CLI »				
	9.3	Licence egov-validationclient-cli	14			
10	Anne	exe	15			
	10.1	Structure du programme	15			
	10.2	Principe de la validation de format	16			

CR 21.02.2024 Page 1/16

1 Préambule

KOST-Val est une application java pour la validation de la configuration et du contenu des fichiers PDF/A, JP2, JPEG, TIFF, PNG, FLAC, WAVE, MP3, MKV, MP4, XML, SIARD 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 9.

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.

2 Configurations système requises

- Microsoft Windows 64bit
- Au moins 512 Mo de RAM
- Au moins 20 Go d'espace disque

3 Problèmes en suspens / Commentaires / Souhaits de modifications

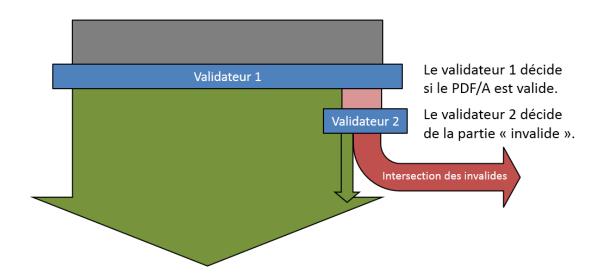
Les points ouverts (erreurs, bogues, souhaits de modifications) sont visibles sur la plate-forme de développement GitHub (https://github.com/KOST-CECO/KOST-Val/issues) et peuvent être rapportés à kost-val@kost-ceco.ch.

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

CR 21.02.2024 Page 2/16

4 Introduction à la 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.¹



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.

Pour la double validation, « 3-Heights™ PDF/A Validator » de PDF-Tools et « pdfaPilot » de callas sont utilisés. Si seulement un validateur est activé, 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.

-

¹ 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 « pdfaPilot » de callas répondent à ces exigences élevées.

5 Installation de KOST-Val

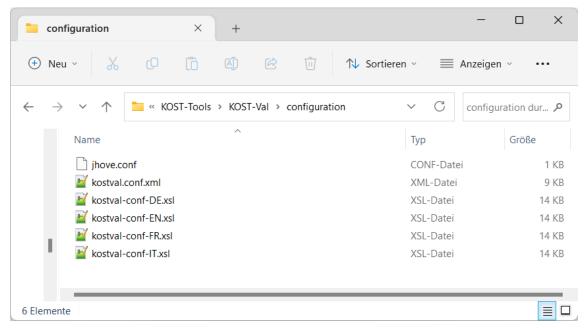
1 KOST-Val (version 2.1.3.0 et plus récente) est disponible que dans le paquet d'installation 64bit KOST-Tools.msi².

https://github.com/KOST-CECO/KOST-Val/releases/latest

Après le téléchargement de KOST-Tools, le paquet d'installation doit être exécuté avec des droits d'administration.

KOST-Val est ensuite disponible dans le menu de démarrage sous KOST-Tools.

6 Configuration de KOST-Val



Le fichier « jhove.conf » est enregistrés dans le dossier « configuration » et n'a pas besoin d'être ajusté. « jhove.conf » est requis pour la validation de Jhove intégré dans KOST-Val.

Le fichier de configuration « kostval.conf.xml » ainsi que les quatre stylesheets sont copié dans le répertoire « USERHOME/.kost-val_2x/configuration » s'ils ne sont pas corrects ou actuels. Toutes les configurations du KOST-Val peuvent être effectuées via l'interface graphique (GUI).

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

Le fichier de configuration « kostval.conf.xml » est divisé en différentes parties. Ci-après une brève description des composants.

.

² Des instructions plus détaillées sur l'installation et son étendue peuvent être consultées dans le manuel KOST-Tools.

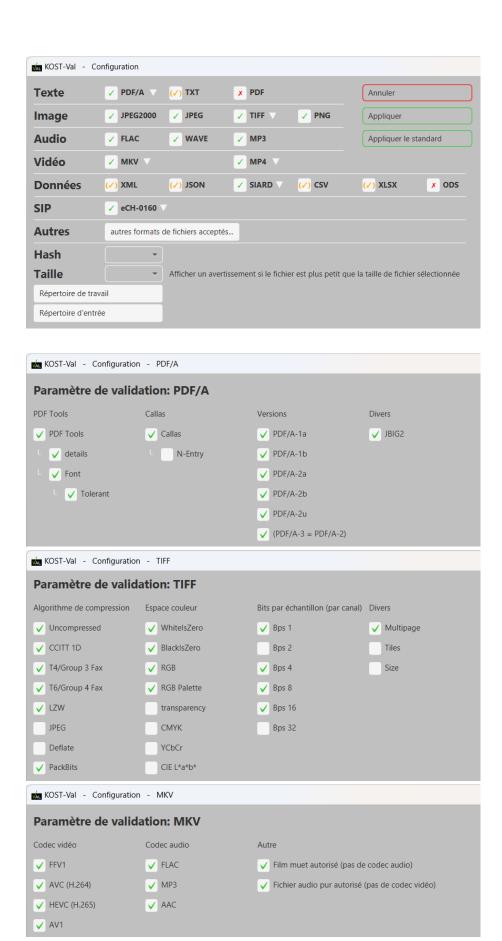


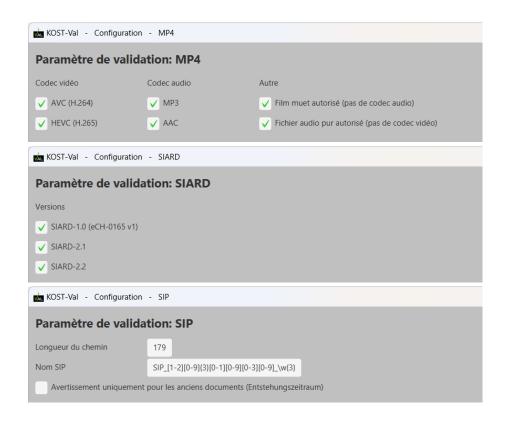
Légende :

= accepté et validé (✓) = accepté x = non accepté

DDE/A: Accordation at validation [/]	✓	
PDF/A : Acceptation et validation [✓] Validation PDF/A avec PDF Tools [yes] :		
	yes	
- PDF Tools également des erreurs détaillées en anglais [yes] :	yes	
- Validation (possibilité de recherche et d'extraction) [tolérant] :	tolerant	
Validation PDF/A avec callas [yes] :	yes	
- Erreur (E) / avertissement (W) si l'entrée N ne correspond pas [W] :	W	
Versions PDF/A autorisées [1A, 1B, 2A, 2B, 2U] :	1A 1B 2A 2B 2U	
Valider le PDF/A-3 vers le PDF/A-2 et générer un avertissement au lieu d'une erreur [yes]:	yes	
Compression JBIG2 autorisée [yes] :	yes	
TXT : Acceptation [(✓)]	(✓)	
PDF: Acceptation [X]	X	
JPEG2000 : Acceptation et validation [✓]	✓	
JPEG : Acceptation et validation [✓]	✓	
TIFF: Acceptation et validation [✓]	/	
Algorithmes de compression autorisés [uncompressed, CCITT 1D, CCITT Group 3, CCITT Group 4, LZW, PackBits] :	uncompressed CCITT 1D CCITT Group 3 CCITT Group 4 LZW PackBits	
Espaces colorimétriques autorisés [white is zero, black is zero, RGB, palette color] :	white is zero black is zero RGB palette color	
Bits autorisés par échantillon [1, 4, 8, 16] :	1 4 8 16	
TIFF multipages autorisés [yes] :	yes	
Structure en carreaux autorisée [no] :	no	
Taille de fichier de 1000MB (~1GB) et plus autorisée [no] :	no	
PNG : Acceptation et validation [✓]	/	
FLAC: Acceptation [/]	/	
WAVE : Acceptation [✓]	/	
MP3 : Acceptation []	<i>'</i>	
	/	
MKV : Acceptation et validation []	FEVALANC HEVC AVA	
- Codec vidéo autorisé [FFV1, AVC, HEVC, AV1] :	FFV1 AVC HEVC AV1	
- Codec audio autorisé [FLAC, MP3, AAC] :	FLAC MP3 AAC	
- Film muet autorisé (pas de codec audio) [Warning]:	Warning	
- Fichier audio pur autorisé (pas de codec vidéo) [Warning]:	Warning	
MP4 : Acceptation et validation [✓]	/	
- Codec vidéo autorisé [AVC, HEVC] :	AVC HEVC	
- Codec audio autorisé [MP3, AAC] :	MP3 AAC	
- Film muet autorisé (pas de codec audio) [Warning]:	Warning	
- Fichier audio pur autorisé (pas de codec vidéo) [Warning]:	Warning	
XML : Acceptation et validation [(✓)]	(✓)	
JSON : Acceptation [(✓)]	(✓)	
SIARD : Acceptation et validation [✓]	✓	
Versions SIARD autorisées [1.0, 2.1, 2.2] :	1.0 2.1 2.2	
CSV : Acceptation [(✓)]	(✓)	
XLSX : Acceptation [(✓)]	(/)	
ODS : Acceptation [X]	×	
SIP : Validation [✓]:	/	
Nombre maximal de caractères autorisés dans les longueurs de chemin [179] :	179	
Spécifications pour la structure du nom SID [SID [1-2][0-0][3][0-1][0-0][0.3]	SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_\w{3}	
Avertissement uniquement pour les anciens documents (Entstehungszeitraum)	no	
[no]: Autres formats de fichiers acceptés [WARC, HTML, DWG]:	HTML WARC DWG	
Calculer et afficher la valeur de hachage des fichiers. Vide signifie pas de	TITLE WARE DWG	
calcul ni de sortie [] : Afficher un avertissement pour les petits fichiers. Vide signifie pas		
d'avertissement [] : Répertoire de travail [] :		
Répertoire d'entrée [] :		
repertone a charee [] .		

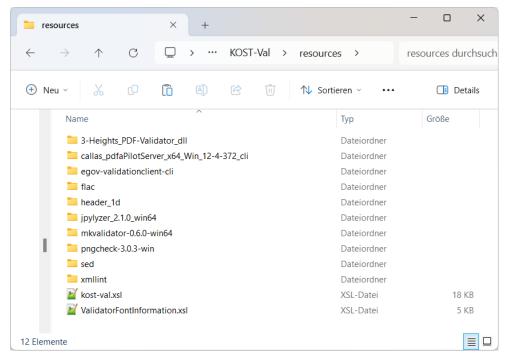
CR 21.02.2024 Page 5/16





7 Ressources de KOST-Val

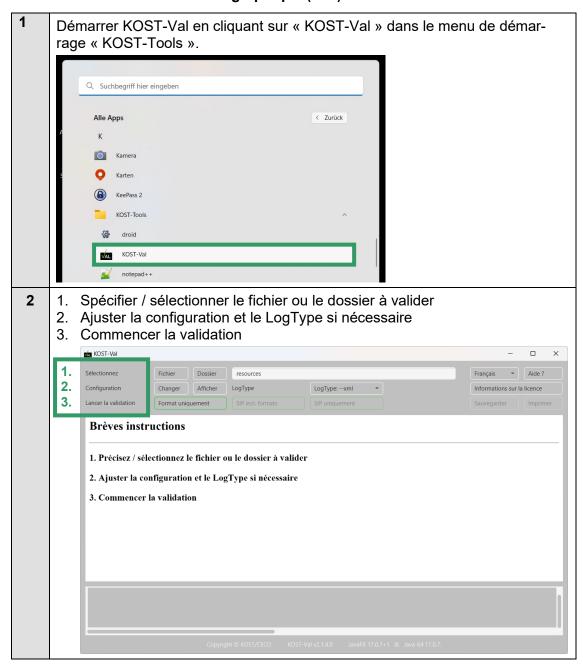
Toutes les ressources de KOST-Val sont stockées dans le sous-dossier « ressources ».



8 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.

8.1 KOST-Val avec l'interface graphique (GUI)



CR 21.02.2024 Page 9/16

4. Attendre que la validation soit terminée. Le dossier en cours est indiqué dans le champ inférieur. KOST-Val Sélectionnez Changer Afficher LogType Lancer la validation La validation du format est lancée. 4. Veuillez patienter ... 3 A la fin de la validation, le fichier journal est affiché. Il contient des informations détaillées supplémentaires sur les différentes étapes de validation non valables, en particulier l'étape de validation concernée et l'erreur correspondante. Si vous le souhaitez, le fichier journal de KOST Val peut être enregistré ou imprimé. KOST-Val - 🗆 X Fichier Dossier resources Français ▼ Aide ? Changer Afficher LogType LogType: --xml ▼ Informations sur la licence **KOST-Val** Validation: PDF/A, JP2, JPEG, TIFF, PNG, SIARD Reconnaissance: TXT, FLAC, WAVE, MP3, MKV, MP4, XML, JSON, CSV, XLSX, HTML, WARC, DWG, Parmi les 11542 fichiers 553 (4.791%) sont valides, 40 (0.347%) invalides et 10949 (94.862%) n`ont pas ete acceptees. Invalid: Validation: TIFF -> C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val\resources\ExifTool-12.62\t\images\GeoTiff.tif B) Jhove La validation de Jhove a echoue. Status: Well-Formed, but not valid B) Jhove - Jhove ErrorMessage: Projection value out of range: 16017

CR 21.02.2024 Page 10/16

8.2 Démarrer la validation manuellement

Ouvrir la fenêtre d'exécution et aller dans le répertoire de travail (cd "C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val") 3 . C:\Users>cd "C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val" C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val> 2 Appeler le programme KOST-Val (en séparant tous les composants d'entrée par des espaces). ..\Liberica JRE\bin\java.exe -jar cmd KOST-Val.jar --format resources --fr --xml C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val>..\Liberica_JRE\bin\java.exe -jar cmd_KOST-Val.jar --format resources Notes: L'entrée java -jar est seulement possible si le Java Runtime Environment (JRE) souhaité est la version standard. Si nécessaire, la valeur 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 » (java -Xmx1024m -Xss128m -jar). Quand un élément d'entrée contient des espaces, celui-ci doit être inscrit entre guillemets. Il est également possible d'appeler KOST-Val depuis tout autre emplacement. Cependant, cela nécessite l'utilisation de liens absolus. Commande pour KOST-Val: java -jar cmdKOST-Val.jar-lien val-type lien et fichier cmdKOST-Val.jar
type de validation:
--format (Validation du format uniquement) ou
--sip (Validation du SIP incl. formats)
--onlysip (validation du SIP uniquement)
lien et nom complet du fichier ou dossier a valider
langue: --de ou --fr ou --it ou --en (facultatif)
type de journal de la sortie (facultatif):
--xml (log standard au format xml)
--min (Connexion au format xml avec seulement les donnees valides ou non)
--max (option xml + verbose si vous souhaitez conserver les rapports originaux) 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:\Program Files\KOST-CECO\KOST-Tools\KOST-Val>) apparaît. Validation du format effectuee. -> Parmi les 11542 fichiers 553 (4.791%) sont valides, 40 (0.347%) invalides et 10949 (94.862%) n'ont pas ete accep tees.
-> C:\Users\clair\.kost-val_2x\logs\resources.kost-val.log.xml C:\Program Files\KOST-CECO\KOST-Tools\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

-

³ Le disque est par exemple changé avec c:.

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	6.27.2.1	Voir chapitre 9.1
Apache Commons http://commons.apache.org/		Apache License 2.0
- commons-logging-1.2.jar	1.2 2.13	
- commons-io-2.13.jar	2.13	
Apache Xerces2 http://xerces.apache.org/	2.12.2	Apache License 2.0
BadPeggy http://coderslagoon.com/	2.0	GPL v3 License
Jdom http://www.jdom.org/	2.0.6.1	jdom License
Jhove https://jhove.openpreservation.org/	1.28	LGPL v2.1 License
Spring Framework API		
https://spring.io/projects/spring-framework	5.3.19	Apache License 2.0
zip64 <u>http://sourceforge.net/projects/zip64file/</u>	1.02	GPL v2+ License

KOST-Val utilise les composants inchangés suivants d'autres fabricants qui sont livrés avec KOST-Val:

Programme tie	ers/Composants tiers	Version	Licence
egov-validatio	nclient-cli https://www.bit.admin.ch/	1.0.10	Voir chapitre 9.3
flac	https://xiph.org/flac	1.4.3	BSD License
Jpylyzer	http://jpylyzer.openpreservation.org/	2.1.0	LGPL v3.0 License
mkvalidator	https://www.matroska.org/	0.6.0	BSD License
pdfaPilot CLI	https://www.callassoftware.com	12.4.372	Voir chapitre 9.2
pngcheck http	o://libpng.org/pub/png/apps/pngcheck	3.0.3	GPL v2 License
GNU sed	https://www.gnu.org/software/sed	4.4	GPL v3+ License
Xmllint	https://xmllint.com/	20630	MIT 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.

9.1 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 d'accord individuel suivantes aux Conditions Générales de licence avec PDF Tools :

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.

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'usager 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. L'activation de cette licence se fait avec le « LicenseManager.exe », qui existe déjà dans « KOST-Val\resources\3-Heights PDF-Validator dll ».

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

CR 21.02.2024 Page 13/16

9.2 Licence « pdfaPilot CLI »

Pour l'utilisation de la version limitée de « pdfaPilot CLI » de callas, le CECO a convenu des conditions d'accord individuel suivantes aux Conditions Générales de licence avec callas :

2. Individuelle Vereinbarung

Dieses Vertragsverhältnis regelt die Lizenz zwischen der callas software als Lizenzgeber und der KOST als Lizenznehmer gemäss nachfolgenden Spezialbestimmungen:

- callas software erteilt für die KOST eine kostenfreie Lizenz für callas pdfaPilot CLI für Windows zur innerbetrieblichen Nutzung und zur Integration in ihren eigenen Validator "KOST-Val".
- Die Lizenz schliesst die Distribution von KOST-Val an "Anwender" (Gedächtnisinstitutionen, Archive oder Bibliotheken und deren Zulieferer) ein.
- Für die Distribution von KOST-Val an diese Anwender ist die KOST zuständig und darf KOST-Val auch gratis und ohne Registrierung an diese abgeben.
- Die Lizenz ist zeitlich unbegrenzt, jedoch bezüglich Durchsatz pro Installation begrenzt auf 72'000 Seiten pro Jahr.
- Die KOST darf den eigenen Quellcode von KOST-Val Open Source publizieren. callas pdfaPilot CLI ist hiervon ausgenommen.
- First Level Support der Anwender erfolgt durch die KOST. Second Level Support leistet callas software gegenüber der KOST.

Les points suivants sont déterminants pour l'utilisateur :

- La licence comprend la distribution de KOST-Val aux « utilisateurs » (institutions patrimoniales, archives, bibliothèques, et leurs fournisseurs).
- La licence est illimitée dans le temps, mais le débit par installation est limité à 72'000 pages par an.
- Le CECO peut publier leur propre code de KOST-Val open source mais pdfaPilot CLI est exclue.
- Le support de niveau 1 aux usagers est effectué par le CECO. Le CECO transmet à callas software les cas relevant du support de niveau 2.

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

9.3 Licence egov-validationclient-cli

Pour la validation de signatures électroniques dans des fichiers PDF / PDF/A, une licence pour le validateur de signature discrète de la Confédération est nécessaire. Cette licence n'est pas incluse dans le CECO-Val. Plus d'informations sont disponibles sur https://www.bit.admin.ch/bit/fr/home/themen/elektronische-signatur/anzeigen-und-validieren/der-diskrete-validator/validator-web-service.html.

Aucune validation de signature n'est encore effectuée dans CECO-Val 2.2.0.0. Jusqu'à présent, seule la reconnaissance a été programmée et ne nécessite pas de licence.

10 Annexe

10.1 Structure du programme

KOST-Val a été développé selon les exigences suivantes :

Exigences fonctionnelles:

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 du module de validation, les résultats valides ou invalides ainsi que le chemin vers le fichier. Toutes les informations additionnelles sont spécifiées dans le fichier log.

CR 21.02.2024 Page 15/16

10.2 Principe de la validation de format

