



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 à la double validation PDF/A.....	4
5	Installation de KOST-Val.....	5
6	Configuration de KOST-Val.....	5
6.1	Composants du fichier de configuration « kostval.conf.xml ».....	5
6.2	« KaD_SignatureFile_V72.xml » et « jhove.conf » Fehler! Textmarke nicht definiert.	
7	Ressources de KOST-Val	8
8	Exécuter la validation	9
8.1	KOST-Val avec l'interface graphique (GUI)	9
8.2	Démarrer la validation manuellement	11
9	Description des étapes de validation.....	13
9.1	TIFF	13
9.2	SIARD	14
9.3	PDF/A	15
9.1	JP2.....	16
9.2	JPEG.....	16
9.3	PNG	17
9.4	SIP	18
10	Droits d'auteur.....	19
10.1	Licence « 3-Heights™ PDF/A Validator API »	20
10.2	Licence « pdfaPilot CLI ».....	21
11	Annexe.....	22
11.1	Structure du programme.....	22
11.2	Principe de la validation de format.....	24

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*), JPEG, PNG (*Portable Network Graphics*), XML (*Extensible Markup Language*) 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 10.

KOST-Val remplit les exigences décrites ci-après.

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 « pdfaPilot » de callas 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 des versions limitée¹ de « 3-Heights™ PDF/A Validator » de PDF-Tools et « pdfaPilot » de callas.

Dans le module J il peut être défini si la compression JBIG2² est acceptée ou non. Le module K contrôle, selon la configuration, si les fonts peuvent être recherchées et extraites ou non.

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

Validation JPEG : KOST-Val lit un fichier JPEG (ISO 10918-1) et valide avec l'aide de BadPeggy la structure et le contenu³.

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

Validation PNG : KOST-Val lit un fichier PNG (ISO 15948) et valide la structure et le contenu avec l'aide de pngcheck.

Validation XML : KOST-Val lit un fichier XML et valide la structure et le contenu avec l'aide de xmllint et XSD.

Validation SIARD : KOST-Val lit un fichier SIARD (eCH-0165⁴ v1 et SIARD v2.1⁵) et valide la structure et le contenu.

¹ La restriction est principalement liée à la capacité maximale de 72'000 pages par an. Plus d'informations sur les licences sont dans chapitre 10. 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 LicenseManager (resources\3-Heights_PDF-Validator_dll) respectivement par acquérant « pdfaPilot » de callas.

² Au sujet du problème de la compression JBIG2 voir http://kost-ceco.ch/cms/index.php?jbig2-compression_fr. Le Groupe d'experts du preservation planning PPEG recommande pour le moment de ne pas utiliser la compression JBIG2 lors de la création de fichiers PDF.

³ Le message d'erreur « Not a JPEG file » est raffiné par KOST-Val et sortie différencié.

⁴ La spécification peut être téléchargée à partir du site eCH : <https://www.ech.ch/ech/ech-0165>.

⁵ La spécification peut être téléchargée à partir du site CECO : https://kost-ceco.ch/cms/index.php?siard_de.

Validation SIP : KOST-Val lit un SIP (eCH-0160⁶ v1.0, v1.1 et v1.2) 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.

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

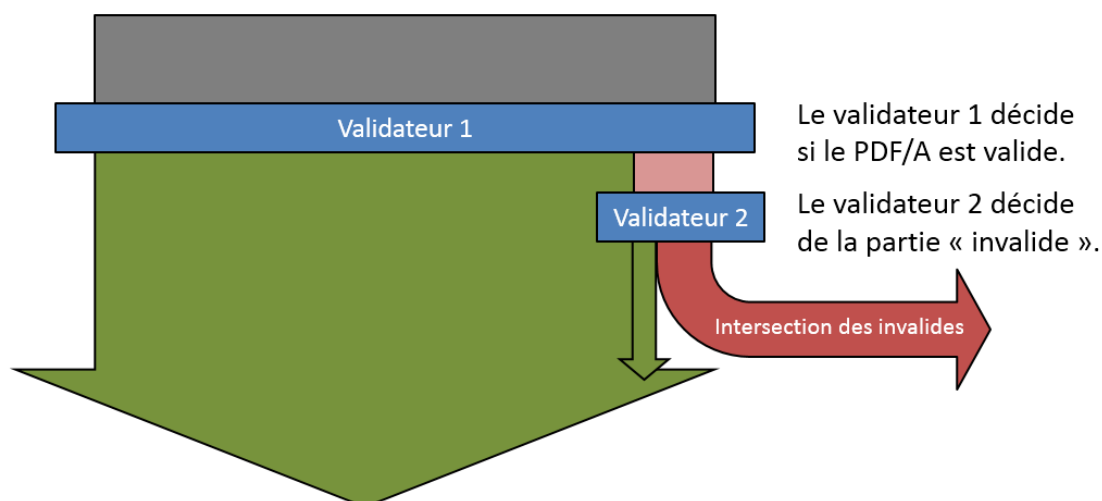
Les points ouverts (erreurs, bogues, souhaits de modifications) sont visibles sur la plate-forme de développement GitHub (<https://github.com/KOST-CECO/KOST-Val/issues>) et peuvent être rapportés à kost-val@kost-ceco.ch.

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

⁶ La spécification peut être téléchargée à partir du site eCH : <https://www.ech.ch/fr/eCH/eCH-0160>.

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.⁷



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.

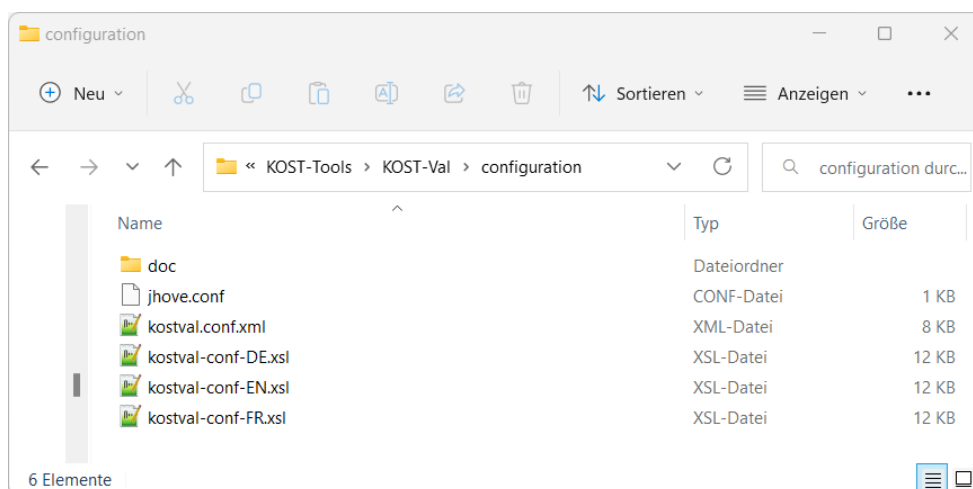
Le constat que même des validateurs PDF/A de qualité peuvent donner des résultats différents est à la base de la conception de la double validation. Cela tient d'une part au fait que la norme PDF/A originelle implique un ensemble d'autres normes et standards qui ne sont pas implémentés jusque dans les 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.

⁷ 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.msi⁸.
<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 enregistré 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 trois 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. La configuration délivrée permet la validation immédiate de PDF/A, JP2, JPEG, TIFF, PNG, XML, SIARD et SIP. En outre, les formats FLAC, WAVE, MP3, MKV, MP4, CSV, XLSX, WARC, HTML et DWG sont acceptés dans les dossiers et les SIP. Ci-après une brève description des composants.

⁸ Des instructions plus détaillées sur l'installation et son étendue peuvent être consultées dans le manuel KOST-Tools.

KOST-Val Configuration

Légende :

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

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] :	tolérant
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
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, T4/Group 3 Fax, T6/Group 4 Fax, LZW, PackBits) :	Uncompressed CCITT 1D T4/Group 3 Fax T6/Group 4 Fax LZW PackBits
Espaces colorimétriques autorisés [WhitIsZero, BlackIsZero, RGB, RGB Palette] :	WhitIsZero BlackIsZero RGB RGB Palette
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 [(✓)]	(✓)
MKV (FFV1-FLAC) : Acceptation [(✓)]	(✓)
MP4 (AVC-MP3) : Acceptation [(✓)]	(✓)
XML : Acceptation et validation [✓]	✓
SIARD : Acceptation et validation [✓]	✓
Versions SIARD autorisées [1.0, 2.1] :	1.0 2.1
CSV : Acceptation [(✓)]	(✓)
XLSX : Acceptation [(✓)]	(✓)
ODS : Acceptation [x]	x
SIP : Validation [✓]	✓
Nombre maximal de caractères autorisés dans les longueurs de chemin [179] :	179
Spécifications pour la structure du nom SIP [SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_w{3}] :	SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_w{3}
Autres formats de fichiers acceptés [WARC, HTML, DWG] :	WARC HTML DWG
Répertoire de travail [] :	
Répertoire d'entrée [] :	

KOST-Val - Configuration

Texte

☒ PDF/A *
 ☐ TXT
 ☐ PDF

Image

☒ JPEG2000
 ☒ JPEG
 ☒ TIFF *
 ☒ PNG

Audio

☐ FLAC
 ☐ WAVE
 ☐ MP3

Vidéo

☐ MKV (FFV1-FLAC)
 ☐ MP4 (AVC-MP3)

Données

☒ XML
 ☒ SIARD *
 ☐ CSV
 ☐ XLSX
 ☐ ODS

SIP

☒ eCH-0160 *

Autres

KOST-Val - Configuration - PDF/A

Paramètre de validation: PDF/A

PDF Tools

☒ PDF Tools
 ☐ details
 ☐ Font
 ☐ Tolerant

Callas

☒ Callas
 ☐ N-Entry

Versions

☒ PDF/A-1a
 ☒ PDF/A-1b
 ☒ PDF/A-2a
 ☒ PDF/A-2b
 ☒ PDF/A-2u

Divers

☒ JBIG2

Paramètre de validation: TIFF

Algorithme de compression	Espace couleur	Bits par échantillon (par canal)	Divers
<input checked="" type="checkbox"/> Uncompressed	<input checked="" type="checkbox"/> WhitelsZero	<input checked="" type="checkbox"/> Bps 1	<input checked="" type="checkbox"/> Multipage
<input checked="" type="checkbox"/> CCITT 1D	<input checked="" type="checkbox"/> BlacksZero	<input type="checkbox"/> Bps 2	<input type="checkbox"/> Tiles
<input checked="" type="checkbox"/> T4/Group 3 Fax	<input checked="" type="checkbox"/> RGB	<input checked="" type="checkbox"/> Bps 4	<input type="checkbox"/> Size
<input checked="" type="checkbox"/> T6/Group 4 Fax	<input checked="" type="checkbox"/> RGB Palette	<input checked="" type="checkbox"/> Bps 8	
<input checked="" type="checkbox"/> LZW	<input type="checkbox"/> transparency	<input checked="" type="checkbox"/> Bps 16	
<input type="checkbox"/> JPEG	<input type="checkbox"/> CMYK	<input type="checkbox"/> Bps 32	
<input type="checkbox"/> Deflate	<input type="checkbox"/> YCbCr		
<input checked="" type="checkbox"/> PackBits	<input type="checkbox"/> CIE L*a*b*		

Paramètre de validation: SIARD

Versions

- ☒ SIARD-1.0 (eCH-0165 v1)
- ☒ SIARD-2.1

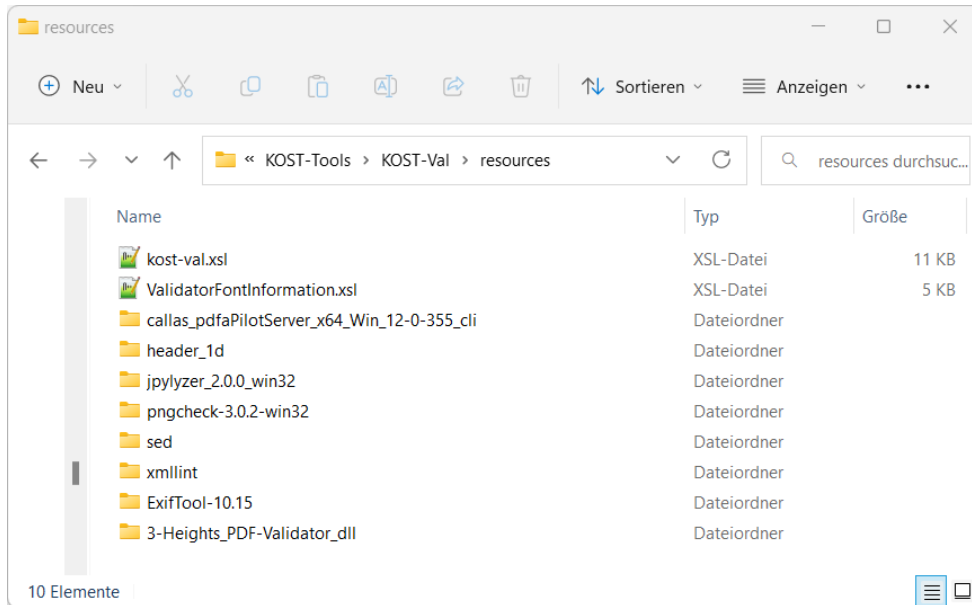
Paramètre de validation: SIP

Longueur du chemin

Nom SIP

7 Ressourcen de KOST-Val

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



8 Exécuter la validation



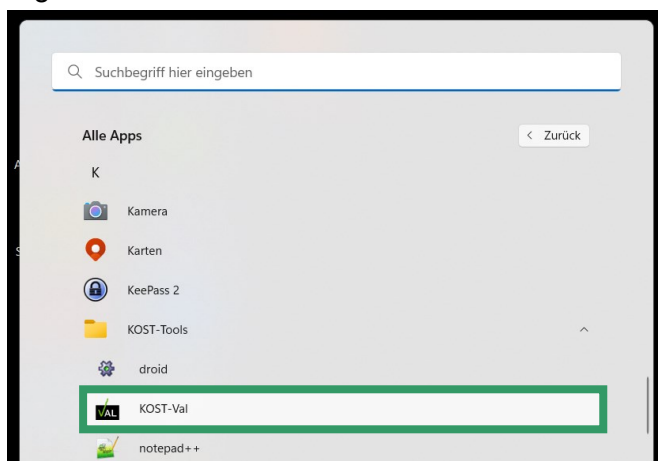
KOST-Val n'est pas thread-safe!

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)

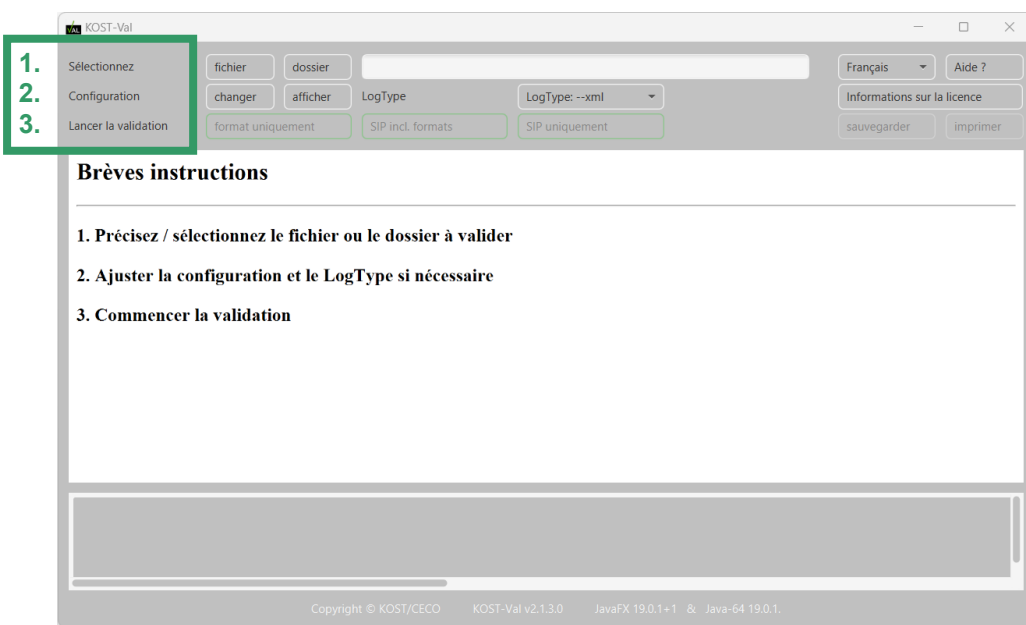
1

Démarrer KOST-Val en cliquant sur « KOST-Val » dans le menu de démarrage « KOST-Tools ».

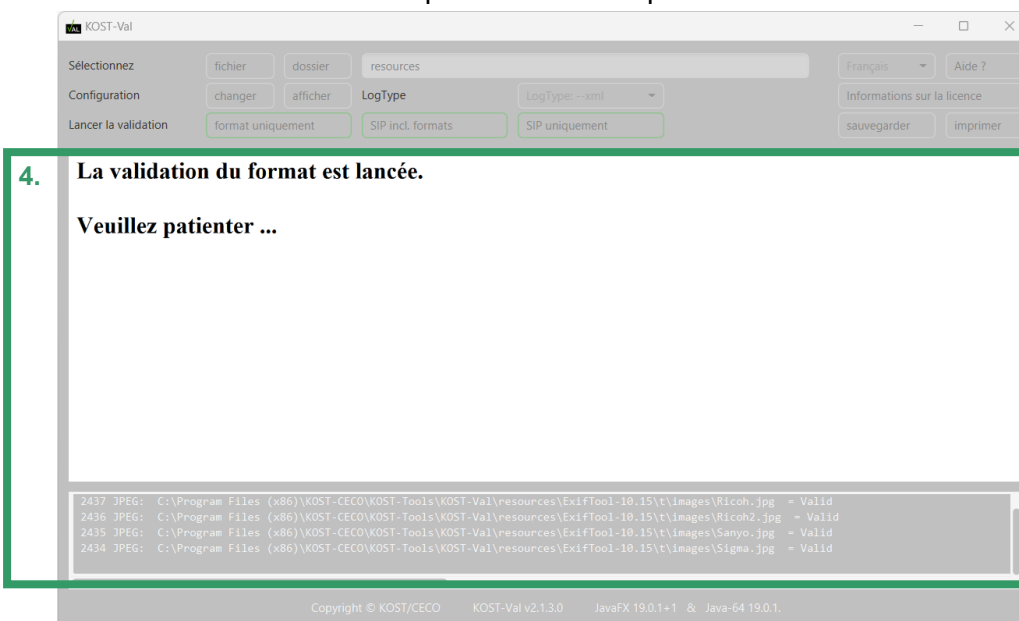


2

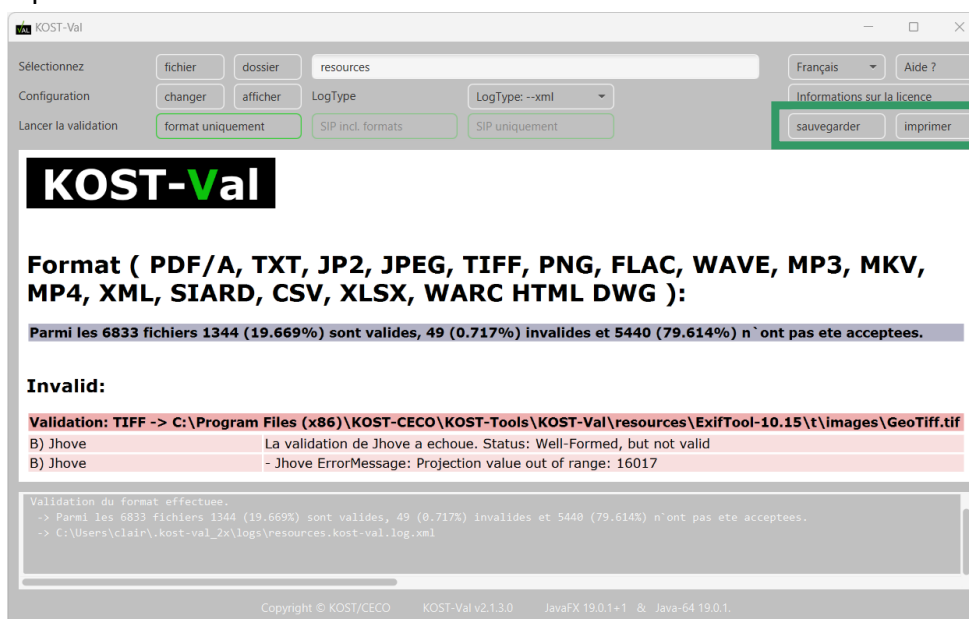
1. Spécifier / sélectionner le fichier ou le dossier à valider
2. Ajuster la configuration et le LogType si nécessaire
3. Commencer la validation



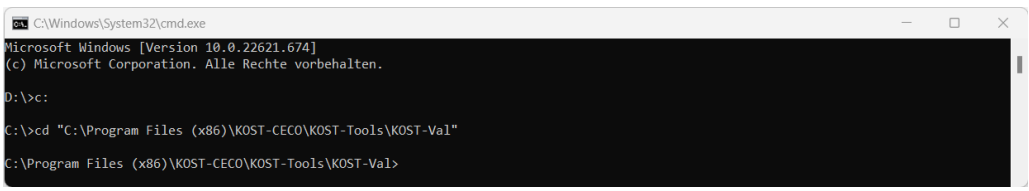
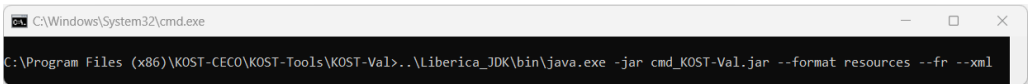
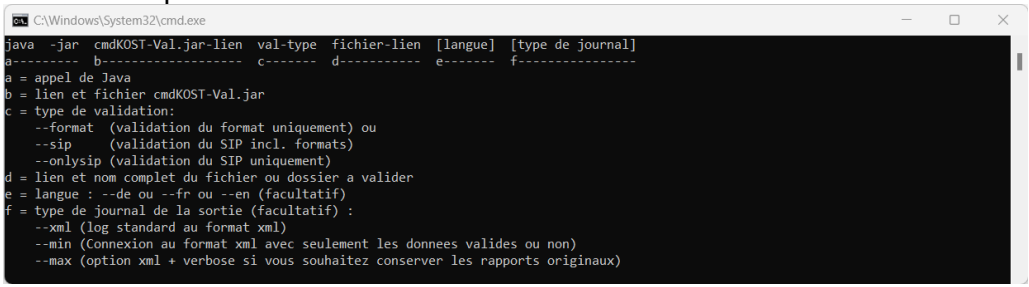
4. Attendre que la validation soit terminée.
Le dossier en cours est indiqué dans le champ inférieur.



- 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.



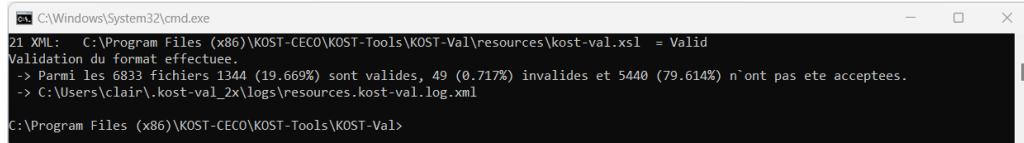
8.2 Démarrer la validation manuellement

1	<p>Ouvrir la fenêtre d'exécution et aller dans le répertoire de travail (<code>cd "C:\Program Files (x86)\KOST-CECO\KOST-Tools\KOST-Val"</code>)⁹.</p> 
2	<p>Appeler le programme KOST-Val (en séparant tous les composants d'entrée par des espaces).</p> <pre>..\Liberica_JDK\bin\java.exe -jar cmd_KOST-Val.jar --format resources --fr --xml</pre>  <p><u>Notes :</u></p> <p>L'entrée <code>java -jar</code> est seulement possible si le Java Runtime Environment (JRE) souhaité est la version standard.</p> <p>Si nécessaire, la valeur de la mémoire virtuelle de Java peut être rapidement adaptée. <code>-Xmx</code> doit être adaptée sur « Out of Memory » et <code>-Xss</code> à des erreurs « Stack Overflow » (<code>java -Xmx1024m -Xss128m -jar</code>).</p> <p>Quand un élément d'entrée contient des espaces, celui-ci doit être inscrit entre guillemets.</p> <p>Il est également possible d'appeler KOST-Val depuis tout autre emplacement. Cependant, cela nécessite l'utilisation de liens absolus.</p> <p>Commande pour KOST-Val:</p> 

⁹ Le disque est par exemple changé avec `c :`.

3

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 (x86)\KOST-CECO\KOST-Tools\KOST-Val>) apparaît.



```
C:\Windows\System32\cmd.exe
Z1 XML: C:\Program Files (x86)\KOST-CECO\KOST-Tools\KOST-Val\resources\kost-val.xml = Valid
Validation du format effectuee.
-> Parmi les 6833 fichiers 1344 (19.669%) sont valides, 49 (0.717%) invalides et 5440 (79.614%) n'ont pas ete acceptees.
-> C:\Users\clair\.kost-val_2x\logs\resources.kost-val.log.xml
C:\Program Files (x86)\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

9 Description des étapes de validation

9.1 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 ?
N'existe il pas du « Planar 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 ?

9.2 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? Furent xyz et abc rempli et non laissé sur (...)?
- W avertissement**
Furent <dataOwner> et <dataOriginTimespan> rempli et non laissé sur (...)?

9.3 PDF/A

Les messages du rapport initial de callas et les messages d'erreur détaillés de PDF Tools sont attribués aux modules basés sur des mots. Les messages d'erreur (« Message ») sont pris en charge 1:1. Lors de la validation avec PDF Tools, les catégories sont attribuées aux modules, traduites, et le texte est affiché.

A général

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

B structure

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

C graphique

La validation graphique (par exemple, espace de couleur valide) est-elle réussie ? Le nombre de composants dans l'entrée N du PDF/A OutputIntent correspond-il au profil ICC ?¹⁰

D polices

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

E transparence

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

F annotations

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

G 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 métadonnées

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

I disponibilité

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

J contrôle JBIG2 (configurable)

Le fichier PDF/A contient-il la compression JBIG2 ? ¹¹

K validation des polices (configurable)

Les polices contenues sont-elles consultables et extractibles ?

¹⁰ A cet effet, on effectue un test supplémentaire avec « pdfaPilot » de callas, qui émet un avertissement ou une erreur en fonction de la configuration (voir 6.1). Le « 3-Heights™ PDF/A Validator » de PDF Tools, effectuée cette vérification en standard, mais elle est critiquée dans la iCategory_1, qui est émise en raison de la composition sous « A) validation générale ».

¹¹ Au sujet du problème de la compression JBIG2 voir http://kost-ceco.ch/cms/index.php?jbig2-compression_fr. Le Groupe d'experts du preservation planning PPEG recommande pour le moment de ne pas utiliser la compression JBIG2 lors de la création de fichiers PDF.

9.1 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 ?

9.2 JPEG

Les messages d'erreur de BadPeggy sont attribués aux modules et traduits (original en anglais). Ce chapitre décrit plus en détail les étapes de validation de JPEG.

A Détection et BadPeggy

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

B Données corrompues

BadPeggy génère-il un message d'erreur, ce qui indique des données corrompues°?

C Structure de fichier invalide

BadPeggy génère-il un message d'erreur, ce qui indique une structure de fichier non valide°?

D Autres problèmes

BadPeggy génère-il un message d'erreur, qui n'a pas encore été attribué et traduit°?

9.3 PNG

Les messages d'erreur de pngcheck sont affectés aux modules et traduits en français (original en anglais). Pngcheck ne produit qu'un seul message d'erreur à la fois. Les différentes étapes de validation PNG sont décrites plus en détail ci-dessous.

A Reconnaissance et pngcheck

Le fichier PNG est-il reconnu comme PNG ? La validation du pngcheck a-t-elle été réussie ?

B Structure

Le fichier PNG est-il correctement structuré [début de l'image (IHDR) et fin de l'image (IEND)] ?

C Couleurs

Les couleurs sont-elles correctement définies dans le fichier PNG [palette de couleurs (PLTE), couleur de fond (bKGD), coordonnées de chromaticité (cHRM), histogramme des couleurs (hIST), profil de couleurs ICC (iCCP), palette (sPLT), espace de couleurs sRGB (sRGB) et informations de transparence (tRNS)] ?

D Données d'images

Les données d'image sont-elles correctement intégrées dans le fichier PNG [Données d'image (IDAT)] ?

E Métadonnées

Les métadonnées répondent-elles aux exigences [signatures numériques (dSIG), métadonnées Exif (eXIF), texte UTF-8 (iTt), texte (tEXt), temps (tIME) et texte compressé (zTXt)] ?

F Autres

Le fichier PNG contient-il d'autres erreurs [gamma (gAMA), taille de pixel (pHYs), bits significatifs (sBIT), et image stéréo (sTER)] ?

9.4 XML

Les messages d'erreur de xmllint sont attribués aux modules et traduits en français (original en anglais). Les différentes étapes de validation XML sont décrites plus en détail ci-dessous.

A Reconnaissance et xmllint

Le fichier XML est-il reconnu comme XML ? Les fichiers xmllint nécessaires sont-ils disponibles ?

B Structure du XML

Le fichier XML est-il correctement structuré ?

C Validation XML avec XSD

Le fichier XML est-il valide par rapport au fichier XSD interne ?

9.5 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/ ?¹²

1G taille du paquet et nombre de fichiers

Les restrictions concernant la taille du SIP et le nombre de fichiers primaires sont-elles respectées ?

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

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 ?

¹² 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.

10 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	6.23.0.4	Voir chapitre 10.1
Apache Commons http://commons.apache.org/ - commons-logging-1.2.jar - commons-io-2.6.jar	1.2 2.6	Apache License 2.0
Apache Xerces http://xerces.apache.org/	2.7.1	Apache License 2.0
BadPeggy http://coderslagoon.com/	2.0	GPL v3 License
Jdom 2.0.0 http://www.jdom.org/	2.0.0	jdom License
Jhove https://jhove.openpreservation.org/	1.24.1	LGPL v2.1 License
Junit 4.12 http://www.junit.org/	4.12	CPL v1.0
Spring Framework API http://static.springsource.org/spring/docs/5.3.x/api/	5.3.19	Apache License 2.0
zip64 http://sourceforge.net/projects/zip64file/	1.02	GPL v2+ License

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

Programme tiers/Composants tiers	Version	Licence
ExifTool http://www.sno.phy.queensu.ca/~phil/exiftool/	10.15	PERL respective GPL v3.0 License
Jpylyzer http://jpylyzer.openpreservation.org/	2.0.0	LGPL v3.0 License
pdfaPilot CLI https://www.callassoftware.com	12.0.355	Voir chapitre 10.2
pngcheck http://libpng.org/pub/png/apps/pngcheck	3.0.2	GPL v2 License
GNU sed https://www.gnu.org/software/sed	4.4	GPL v3+ License
Xmllint http://xmlsoft.org/xmllint.html/	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.

10.1 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'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. 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.

10.2 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.

11 Annexe

11.1 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
W	avertissement

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 pdfaPilot de callas les points suivants :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	générale
B	structure
C	graphique
D	polices
E	transparence
F	annotations & interactions
G	actions
H	métadonnées
I	disponibilité
J (configurable)	contrôle JBIG2
K (configurable)	validation des polices

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 en cas d'erreur)	détection et Jpylyzer
B	validation des métadonnées
C	validation de l'image
D	validation étendue

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

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	détection et BadPeggy
B	données corremplies
C	structure de fichier invalide
D	autres problèmes

Validation PNG : KOST-Val lit un fichier PNG et valide la structure et le contenu avec l'aide de pngcheck :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	reconnaissance et pngcheck
B	structure
C	couleurs
D	données d'images
E	métadonnées
F	autres

Validation XML : KOST-Val lit un fichier XML et valide la structure et le contenu avec l'aide de xmllint :

Étape de validation	Nom de l'étape
A (abandon en cas d'erreur)	reconnaissance et xmllint
B (abandon en cas d'erreur)	Structure du XML
C	Validation XML avec XSD

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
1g	taille du paquet et nombre de fichiers
2a	données primaires manquantes
2b	données primaires supplémentaires
2c	validation de la somme de contrôle
2d	description des données primaires
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 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.

11.2 Principe de la validation de format

