Ein Gemeinschaftsunternehmen von Schweizer Archiven



Inhalt

1	Vorw	ort				
2	Systemvoraussetzungen					
3	Offene Punkte / Feedback					
4	Einleitung duale PDF/A-Validierung					
5	Installation von KOST-Val					
6	Konfi	4				
	6.1	Bestandteile der Konfigurationsdatei «kostval.conf.xml»	4			
7	Ress	ourcen von KOST-Val	8			
8	Validierung starten					
	8.1	KOST-Val GUI	9			
	8.2	Validierung manuell starten	11			
9	Urheberrecht					
	9.1	3-Heights™ PDF/A Validator API-Lizenz	13			
	9.2	pdfaPilot CLI Lizenz	14			
	9.3	egov-validationclient-cli Lizenz	14			
10	Anha	ıng	15			
	10.1	Programmaufbau	15			
	10.2	Funktionsprinzip Formatvalidierung	16			

1 Vorwort

KOST-Val ist eine java-basierte Anwendung zur Validierung von Aufbau und Inhalt von PDF/A-, JP2-, JPEG-, TIFF-, PNG-, FLAC-, WAVE-, MP3-, MKV-, MP4-, XML- und SIARD-Dateien sowie von sogenannter Submission Information Package (SIP) zur Ablieferung von digitalen Informationen. Diese Anwendung steht unter der GPL3+ Lizenz und wird durch die KOST der Öffentlichkeit quelloffen zur Verfügung gestellt. KOST-Val stützt sich auf unveränderte Komponenten anderer Hersteller, welche direkt im Quellcode von KOST-Val eingebunden sind. Die Benutzer von KOST-Val sind gehalten, die Lizenzbestimmungen all dieser Komponenten zu befolgen. Ausführliche Informationen sind im Kapitel 9 ersichtlich.

Die Resultate (inklusive Meldungen zu Inkonsistenzen oder Fehler) werden pro Schritt ausgegeben und in eine Validierungs-Logdatei geschrieben.

Die einzelnen Validierungsschritte / Prüfungen werden nacheinander ausgeführt. Wo möglich, wird die Validierung auch bei Fehlern weiter fortgesetzt, um die Anzahl von Korrekturzyklen zu reduzieren.

2 Systemvoraussetzungen

- 64bit Microsoft Windows
- Mindestens 512 MB RAM
- Mindestens 20 GB Festplattenspeicher

3 Offene Punkte / Feedback

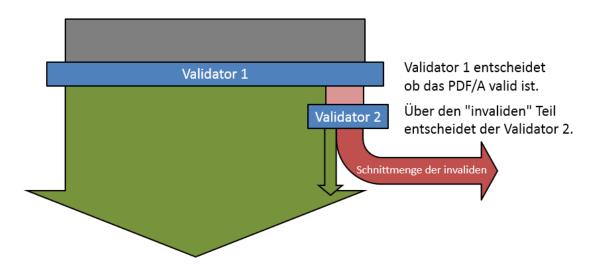
Die offenen Punkte, von Bugs respektive Fehlern bis hin zu Ergänzungswünschen und Fragen, sind auf der Entwicklungsplattform GitHub unter Issues ersichtlich (https://github.com/KOST-CECO/KOST-Val/issues) und können an kost-val@kost-ceco.ch gemeldet werden.

Diese Liste kann und soll durch jedermann erweitert werden und wird durch das Entwicklerteam bearbeitet.

CR 21.02.2024 Seite 2/16

4 Einleitung duale PDF/A-Validierung

Für PDF/A bietet KOST-Val die Möglichkeit einer dualen Validierung. Dabei wird eine PDF/A-Datei zunächst durch einen ersten Validator geprüft. Bei invalidem Resultat folgt eine Prüfung durch einen zweiten Validator. Die PDF/A-Datei gilt als valid, wenn mindestens einer der Validatoren sie als valid identifiziert, und als invalid, wenn beide Validatoren sie als invalid identifizieren.¹



Die duale PDF/A-Validierung darf nur angewendet werden, wenn das Archiv es zulässt, dass potenziell invalide PDF/A-Dateien übernommen werden dürfen. Wenn dies nicht der Fall ist, dann sollte auf die duale PDF/A-Validierung verzichtet werden.

Für die duale Validierung wird sowohl 3-Heights™ PDF/A Validator von PDF-Tools als auch pdfaPilot von callas verwendet. Wenn nur ein Validator eingeschaltet ist, wird automatisch eine einfache Validierung durchgeführt.

Konzeptionelle Grundlage für die duale Validierung ist die Feststellung, dass selbst qualitativ hochstehende PDF/A-Validatoren zu unterschiedlichen Resultaten kommen können. Dies liegt einerseits daran, dass der eigentliche PDF/A-Standard ein Set von anderen Standards einschliesst, welche in den Validatoren nicht zwingend bis in alle Details implementiert sind. Anderseits sind gewisse Vorgaben des Standards so formuliert, dass sie legitimerweise auf verschiedene Arten implementiert werden können. Dass sämtliche relevanten Tools die Spezifikation einheitlich und vollständig implementieren, bleibt vorerst Zukunftsmusik. Deshalb bietet KOST-Val als Zwischenlösung die duale Validierung an.

-

¹ Die duale Validierung kann nur mit qualitativ hochstehenden PDF/A-Validatoren in diesem Sinne durchgeführt werden. Diese hohen Anforderungen erfüllen unter anderem die neusten Versionen von 3-Heights™ PDF/A Validator von PDF-Tools und pdfaPilot von callas.

5 Installation von KOST-Val

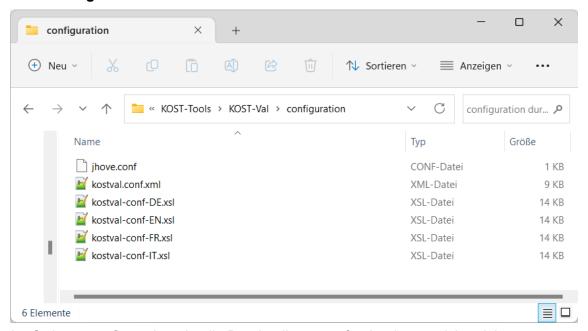
1 KOST-Val (Version 2.1.3.0 und neuer) wird nur noch im 64bit Installationspaket KOST-Tools.msi² angeboten.

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

Nach dem herunterladen von KOST-Tools muss das Installationspaket mit Administrationsrechten ausgeführt werden.

KOST-Val ist im Anschluss im Startmenu unter KOST-Tools vorhanden.

6 Konfiguration von KOST-Val



Im Ordner «configuration» ist die Dateie «jhove.conf» abgelegt, welche nicht angepasst werden muss. «jhove.conf» wird für die interne Jhove-Validierung benötigt. Die Konfigurationsdatei «kostval.conf.xml» sowie die vier Stylesheets werden, falls nicht korrekt oder aktuell vorhanden, ins Verzeichnis «USERHOME/.kost-val_2x/configuration» kopiert. Sämtliche Konfigurationen des KOST-Val können via GUI vorgenommen werden.

6.1 Bestandteile der Konfigurationsdatei «kostval.conf.xml»

Die Konfigurationsdatei «kostval.conf.xml» ist in verschiedenen Teilen aufgebaut. Nachfolgend werden die Bestandteile kurz beschrieben.

_

² Detailliertere Anleitung zur Installation und deren Umfang kann aus dem KOST-Tools Handbuch entnommen werden.

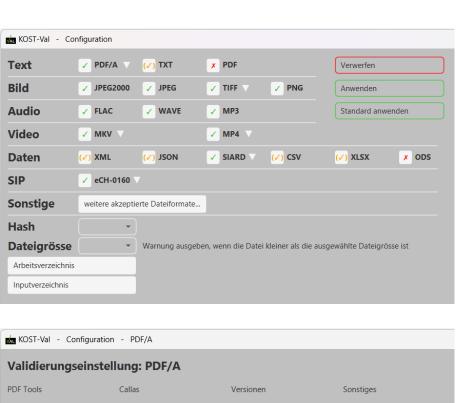


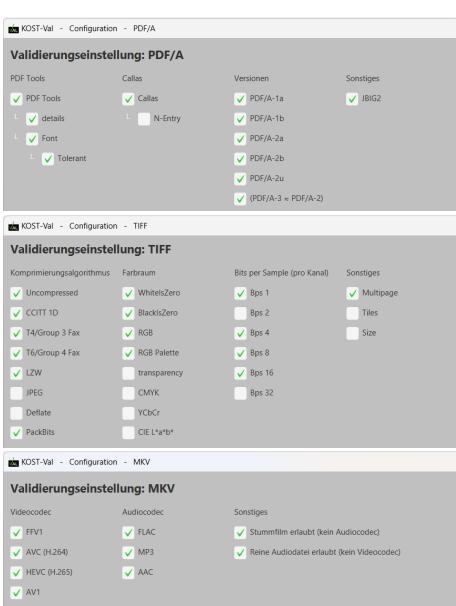
√ = akzeptiert und validieren

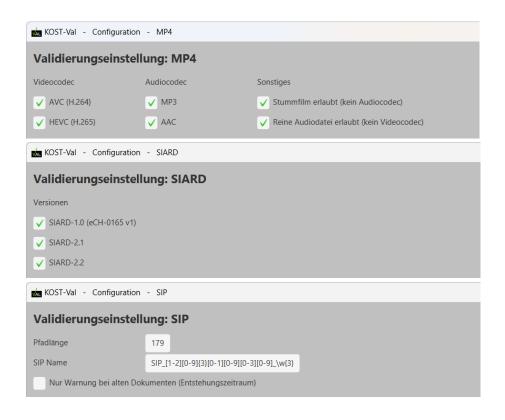
(✓) = akzeptiert x = nicht akzeptiert

DDE/A AL . 13/ H.H. 5 (3)					
PDF/A: Akzeptanz und Validierung [✓]	✓				
PDF/A-Validierung mit PDF Tools [yes]:	yes				
- PDF Tools auch detaillierte Fehler in Englisch [yes]:	yes				
- Validierung (Durchsuch- und Extrahierbarkeit) [tolerant]:	tolerant				
PDF/A-Validierung mit callas [yes]:	yes				
- Fehler (E) / Warnung (W), wenn der N-Eintrag nicht übereinstimmt [W]:	W				
Erlaubte PDF/A Versionen [1A, 1B, 2A, 2B, 2U]:	1A 1B 2A 2B 2U				
PDF/A-3 nach PDF/A-2 validieren und Warnung anstelle eines Fehlers ausgeben [yes]:	yes				
JBIG2-Komprimierung erlaubt [yes]:	yes				
TXT: Akzeptanz [(✓)]	(✓)				
PDF: Akzeptanz [x]	X				
JPEG2000: Akzeptanz und Validierung [✓]	/				
JPEG: Akzeptanz und Validierung [✓]	✓				
TIFF: Akzeptanz und Validierung [✓]	<i>,</i>				
Erlaubten Komprimierungsalgorithmen [uncompressed, CCITT 1D, CCITT Group	uncompressed CCITT 1D CCITT Group 3 CCITT				
3, CCITT Group 4, LZW, PackBits]:	Group 4 LZW PackBits				
Erlaubten Farbräume [white is zero, black is zero, RGB, palette color]:	white is zero black is zero RGB palette color				
Erlaubten Bits per Sample [1, 4, 8, 16]:	1 4 8 16				
Multipage-TIFFs erlaubt [yes]:	yes				
Aufbau in Kacheln erlaubt [no]:	no				
Dateigrössen von 1000MB (~1GB) und grösser erlaubt [no]:	no				
PNG: Akzeptanz und Validierung [🗸]	✓				
FLAC: Akzeptanz [✓]	✓				
WAVE: Akzeptanz [/]	✓				
MP3: Akzeptanz [✓]	✓				
MKV: Akzeptanz und Validierung [✓]	✓				
- Erlaubter Videocodec [FFV1, AVC, HEVC, AV1]:	FFV1 AVC HEVC AV1				
- Erlaubter Audiocodec [FLAC, MP3, AAC]:	FLAC MP3 AAC				
- Stummfilm erlaubt (kein Audiocodec) [Warning]:	Warning				
- Reine Audiodatei erlaubt (kein Videocodec) [Warning]:	Warning				
MP4: Akzeptanz und Validierung [✓]	✓				
- Erlaubter Videocodec [AVC, HEVC]:	AVC HEVC				
- Erlaubter Audiocodec [MP3, AAC]:	MP3 AAC				
- Stummfilm erlaubt (kein Audiocodec) [Warning]:	Warning				
- Reine Audiodatei erlaubt (kein Videocodec) [Warning]:	Warning				
XML: Akzeptanz und Validierung [(✓)]	(✓)				
JSON: Akzeptanz [(✓)]	(v)				
SIARD: Akzeptanz und Validierung [/]	1				
Erlaubte SIARD Versionen [1.0, 2.1, 2.2]:	1.0 2.1 2.2				
CSV: Akzeptanz [(\(\sigma \)]	(/)				
XLSX: Akzeptanz [(✓)]	(v)				
ODS: Akzeptanz [x]	×				
	<i>r</i>				
SIP: Validierung [/]:	170				
Erlaubte maximale Anzahl Zeichen in Pfadlängen [179]:	179				
Vorgaben zum Aufbau des SIP-Namens [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}				
Nur Warnung bei alten Dokumenten (Entstehungszeitraum) [no]:	no				
Weitere akzeptierte Dateiformate [WARC, HTML, DWG]:	HTML WARC DWG				
Hashwert von Dateien berechnen und ausgeben. Leer bedeutet keine Berechnung und Ausgabe []:					
Warnung bei kleinen Dateien ausgeben. Leer bedeutet keine Warnung []:					
Arbeitsverzeichnis []:					
Inputverzeichnis []:					

CR 21.02.2024 Seite 5/16

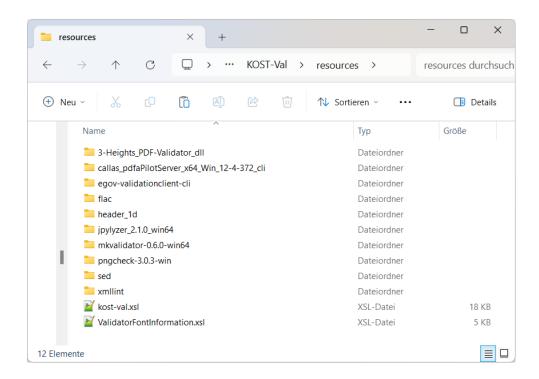






7 Ressourcen von KOST-Val

Sämtliche Ressourcen von KOST-Val sind im Unterordner «resources» abgelegt.



8 Validierung starten

[KOST-Val ist nicht Thread-sicher! Das bedeutet, dass nicht mehrere Instanzen von KOST-Val gleichzeitig ausgeführt werden können, ohne sich gegenseitig zu behindern. Wird KOST-Val gleichzeitig ausgeführt, können Fehler wie z.B. eine fehlende Arbeitskopie vor-

8.1 KOST-Val GUI

kommen.

Starten von KOST-Val mittels Klick auf «KOST-Val» im Startmenu «KOST-Tools».

Suchbegriff hier eingeben

Alle Apps

Kamera

Kamera

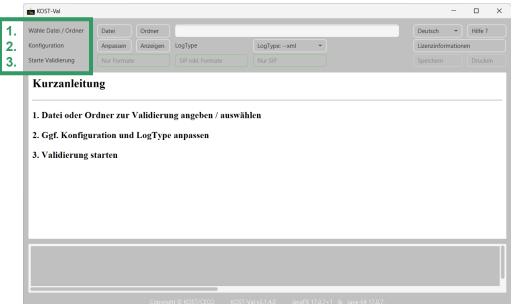
KoST-Tools

KOST-Tools

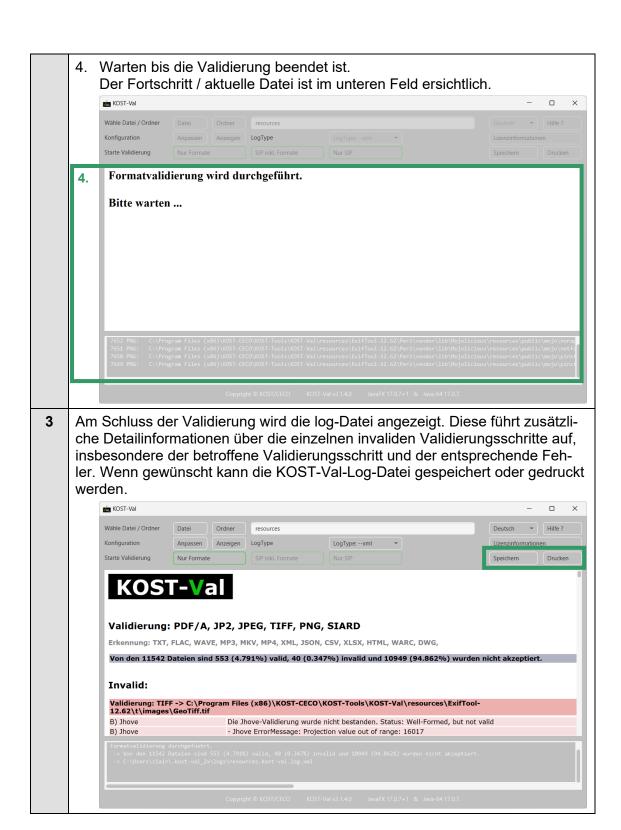
KOST-Tools

- 2 1. Datei oder Ordner zur Validierung angeben / auswählen
 - 2. Ggf. Konfiguration und LogType anpassen
 - 3. Validierung starten

notepad++



CR 21.02.2024 Seite 9/16



CR 21.02.2024 Seite 10/16

8.2 Validierung manuell starten

Eingabeaufforderung öffnen und in das gewünschte Arbeitsverzeichnis wechseln (cd "C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val")³.

2 KOST-Val-Programmaufruf starten (die einzelnen Eingabebestandteile mit Leerzeichen trennen).

```
..\Liberica_JRE\bin\java.exe -jar cmd_KOST-Val.jar --format resources --de --xml
```

```
C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val>..\Liberica_JRE\bin\java.exe -jar cmd_KOST-Val.jar --format resources --de --xml
```

Anmerkungen:

Die Eingabe java -jar ist nur möglich, wenn die gewünschte Java Runtime Environment (JRE) die Standardversion ist.

Bei Bedarf kann die Einstellung des virtuellen Java Memory angepasst werden. -Xmx sollte bei «Out of Memory» und -Xss bei «Stack Overflow» Fehlern angepasst werden (java -Xmx1024m -Xss128m -jar).

Wenn ein Eingabebestandteil Leerzeichen enthält, muss dieser in Anführungsund Schlusszeichen eingegeben werden.

KOST-Val kann auch von einem beliebigen Ort aus aufgerufen werden. Dies bedingt jedoch die Eingabe von absoluten Pfaden.

Aufbau KOST-Val Befehl:

3 Die Datei wurde validiert, sobald "valid" oder "Invalid" im cmd-Fenster erscheint. Der Ordner wurde validiert, sobald die Prompt

(C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val>) erscheint.

```
Formatvalidierung durchgefuehrt.
-> Von den 11542 Dateien sind 553 (4.791%) valid, 40 (0.347%) invalid und 10949 (94.862%) wurden nicht akzeptiert.
-> C:\Users\clair\.kost-val_2x\logs\resources.kost-val.log.xml
C:\Program Files\KOST-CECO\KOST-Tools\KOST-Val>
```

Die detaillierten Resultate sind in der kost-val.log.xml-Datei ersichtlich.

Das Ergebnis der Gesamtvalidierung (korrekte/fehlerhafte Datei) wird ebenfalls ausgegeben und ist im *exit-*Status des Programms sichtbar, so dass die Validierung in eine automatisierte Verarbeitungskette eingebunden werden kann. Der *exit-*Status kann die folgenden Werte annehmen:

- 0 alles OK
- 1 Fehler im Programmaufruf
- 2 Validierung nicht bestanden

³ Das Laufwerk wird z.B. mit c: gewechselt.

9 Urheberrecht

KOST-Val ist eine Entwicklung der KOST. Alle Rechte liegen bei der KOST. KOST-Val wurde im 2012 durch die KOST unter der GNU General Public License v3+ veröffentlicht.

Notice:	This product includes software developed by the Apache Software Foun-
	dation (<u>http://www.apache.org/</u>).

KOST-Val stützt sich auf folgende unveränderte Komponenten anderer Hersteller, welche direkt im Quellcode von KOST-Val eingebunden sind:

Drittprogramm / -Komponente	Version	Lizenz
3-Heights™ PDF/A Validator API		
http://www.pdf-tools.com	6.27.2.1	Siehe Kapitel 9.1
Apache Commons http://commons.apache.org/		Apache License 2.0
- commons-logging-1.2.jar - commons-io-2.13.jar	1.2 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 stützt sich auf folgende unveränderte Komponenten anderer Hersteller, welche mit KOST-Val abgegeben werden:

Drittprogramm	ı / -Komponente	Version	Lizenz
egov-validatio	nclient-cli https://www.bit.admin.ch/	1.0.10	Siehe Kapitel 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	Siehe Kapitel 9.2
pngcheck http	p://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

Die Benutzer von KOST-Val sind gehalten, die Lizenzbestimmungen all dieser Komponenten zu befolgen, welche im Verzeichnis KOST-Val\license vorliegen.

9.1 3-Heights™ PDF/A Validator API-Lizenz

Für die Verwendung der Eingeschränkten Version des 3-Heights™ PDF/A Validator von PDF Tools hat die KOST folgende Individuelle Vereinbarung zu den Allgemeinen Lizenzbedingungen mit PDF Tools vereinbart:

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.

Für die Benutzer sind folgende Punkte massgebend:

- 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).
- 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 Aktivierung dieser Lizenz erfolgt mit dem «LicenseManager.exe», welcher in «KOST-Val\resources\3-Heights PDF-Validator dll» bereits existiert.

Die Benutzer von KOST-Val sind gehalten, diese Lizenzbestimmung zu befolgen.

CR 21.02.2024 Seite 13/16

9.2 pdfaPilot CLI Lizenz

Für die Verwendung der Eingeschränkten Version des pdfaPilot CLI von callas hat die KOST folgende Individuelle Vereinbarung zu den Allgemeinen Lizenzbedingungen mit callas vereinbart:

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.

Für die Benutzer sind folgende Punkte massgebend:

- Die Lizenz schliesst die Distribution von KOST-Val an «Anwender» (Gedächtnisinstitutionen, Archive oder Bibliotheken und deren Zulieferer) ein.
- 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.

Die Benutzer von KOST-Val sind gehalten, diese Lizenzbestimmung zu befolgen.

9.3 egov-validationclient-cli Lizenz

Für die Validierung von elektronischen Signaturen in PDF-/ PDF/A-Dateien wird eine Lizenz für den diskreter Signaturvalidator des Bundes benötigt. Diese Lizenz ist nicht in KOST-Val enthalten. Mehr Informationen sind unter https://www.bit.ad-min.ch/bit/de/home/themen/elektronische-signatur/anzeigen-und-validieren/der-diskrete-validator/validator-web-service.html ersichtlich.

In KOST-Val 2.2.0.0 wird noch keine Signatur-Validierung durchgeführt. Bis anhin ist lediglich die Erkennung programmiert, welche keine Lizenz benötigt.

CR 21.02.2024 Seite 14/16

10 Anhang

10.1 Programmaufbau

KOST-Val wurde nachfolgenden Anforderungen aufgebaut:

Funktionale Anforderungen:

Die Resultate (inklusive Meldungen zu Inkonsistenzen oder Fehler) werden pro Schritt ausgegeben und in eine Validierungs-Logdatei geschrieben.

Das Ergebnis der Gesamtvalidierung (korrekte/fehlerhafte Datei) wird ebenfalls ausgegeben und im *exit*-Status des Programms sichtbar, so dass die Validierung in eine automatisierte Verarbeitungskette eingebunden werden kann. Der *exit*-Status kann die folgenden Werte annehmen:

- 0 alles OK
- 1 Fehler im Programmaufruf
- 2 Validierung nicht bestanden

Die einzelnen Validierungsschritte / Prüfungen werden nacheinander ausgeführt. Wo möglich, wird die Validierung auch bei Fehlern weiter fortgesetzt, um die Anzahl von Korrekturzyklen zu reduzieren.

Nichtfunktionale Anforderungen:

Für besondere Aufgaben werden externe Programme oder entsprechende Java-Frameworks eingesetzt.

Die Anwendung ist modular aufgebaut, damit ohne viel Aufwand ein oder mehrere weitere Validierungsmodule eingebaut werden können.

Die Log-/Programmausgabe erlaubt ein einfaches Auslesen des Ergebnisses der einzelnen Validierung und damit die Verwendung des Tools in einer Prozesskette,

Die Konsolenausgabe begrenzt sich auf die Bezeichnung der Validierungsart, das Gesamtergebnis "valid" oder "invalid" sowie der Pfad zur Datei. Alle zusätzlichen Informationen werden in der Log-Datei aufgeführt.

10.2 Funktionsprinzip Formatvalidierung

