



Anwendungshandbuch

Inhalt

1	Vorwort	2
2	Systemvoraussetzungen	3
3	Offene Punkte / Feedback	3
4	Einleitung duale PDF/A-Validierung	4
5	Installation von KOST-Val	5
6	Konfiguration des KOST-Val	5
6.1	Bestandteile der Konfigurationsdatei "kostval.conf.xml"	5
6.2	"KaD_SignatureFile_V72.xml" und "jhove.conf"	8
7	Ressourcen des KOST-Val	9
8	Validierung starten	10
8.1	Validierung starten mittels "KOST-Val_de.exe"-GUI	10
8.2	Validierung manuell starten	12
9	Validierung auswerten	14
10	Beschreibung der Validierungsschritte	15
10.1	TIFF	15
10.2	SIARD	16
10.3	PDF/A	17
10.4	JP2	18
10.5	JPEG	18
10.6	SIP	19
11	Urheberrecht	20
11.1	3-Heights™ PDF/A Validator API-Lizenz	22
11.2	pdfaPilot CLI Lizenz	23
12	Anhang	24
12.1	Beschreibung der Konfigurationsdatei "kostval.conf.xml"	24
12.2	Programmaufbau	28
12.3	Funktionsprinzip Formatvalidierung	30

1 Vorwort

KOST-Val ist eine java-basierte Anwendung zur Validierung von Aufbau und Inhalt von TIFF-Dateien (*Tagged Image File Format*), SIARD-Dateien (*Software Independent Archiving of Relational Databases*), PDF/A-Dateien, JP2-Dateien (JPEG 2000; *Joint Photographic Experts Group 2000*) JPEG-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 11 ersichtlich.

KOST-Val erfüllt die im Folgenden beschriebenen Anforderungen.

TIFF-Validierung: KOST-Val liest ein TIFF und validiert mit Hilfe von Jhove den Aufbau und Inhalt sowie mit ExifTool die zentralen Eigenschaften wie z.B. Komprimierung, Farbraum und Multipage. Die Eigenschaften sind konfigurierbar.

SIARD-Validierung: KOST-Val liest ein SIARD (eCH-0165¹ v1 und v2.1²) und validiert den Aufbau und Inhalt wie z.B. Struktur, Header- und Content-Validierung.

PDF/A-Validierung: KOST-Val liest ein PDF respektive PDF/A (ISO 19005-1 und 19005-2) und validiert mit Hilfe des 3-Heights™ PDF/A Validator von PDF-Tools oder auch des pdfaPilot von callas den Aufbau und Inhalt der PDF-Datei. KOST-Val gliedert die verschiedenen Fehlermeldungen in Hauptgruppen wie z.B. Schriften, Grafiken und Metadaten. Im Lieferumfang von KOST-Val sind einzig eingeschränkte³ Versionen von 3-Heights™ PDF/A Validator von PDF-Tools sowie pdfaPilot von callas enthalten.

Im Modul J werden je nach Konfiguration zudem die in der PDF-Datei enthaltenen JPEG- und JP2-Bilder extrahiert und validiert⁴. Zusätzlich kann konfiguriert werden, ob die JBIG2-Komprimierung⁵ akzeptiert wird oder nicht.

Im Modul K wird je nach Konfiguration überprüft ob die Schriften durchsuch- und extrahierbar sind oder nicht.

JP2-Validierung: KOST-Val liest ein JP2 (ISO 15444) und validiert mit Hilfe von Jpylyzer den Aufbau und Inhalt.

JPEG-Validierung: KOST-Val liest ein JPEG (ISO 10918-1) und validiert mit Hilfe von BadPeggy den Aufbau und Inhalt⁶.

¹ Die Spezifikation kann von der eCH Website heruntergeladen werden: <https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165&documentVersion=1.0>.

² Die Spezifikation kann von der KOST Website heruntergeladen werden: https://kost-ceco.ch/cms/index.php?siard_de.

³ Die Einschränkung bezieht sich hauptsächlich auf den Durchsatz von bis zu 72'000 Seiten pro Jahr. Mehr Informationen sind in den Lizenzen in Kapitel 11 ersichtlich. Diese Einschränkung kann aufgehoben werden, in dem eine 3-Heights™ PDF/A Validator API Lizenz bei PDF-Tools erworben und mit dem Lizenzmanager aktiviert respektive pdfaPilot bei callas erworben wird.

⁴ ISO 19005 schreibt nicht vor, dass die enthaltenen Bilder valide sein müssen. Diese Validierung der Bilder ist eine archivische Anforderung, insbesondere bei der Verwendung von PDF/A als Bildformat.

⁵ Zur Problematik der JBIG2-Komprimierung siehe http://kost-ceco.ch/cms/index.php?jbig2-compression_de. Die Preservation-Planning-Expertengruppe PPEG der KOST empfiehlt, beim Erstellen von PDF-Dateien vorerst auf die Kompressionsart JBIG2 zu verzichten.

⁶ Die Fehlermeldung „Not a JPEG file“ wird jedoch von KOST-Val weiter ausgewertet und differenziert ausgegeben.

SIP-Validierung: KOST-Val liest ein SIP (eCH-0160⁷ v1 und v1.1) und validiert die Muss-Punkte aus der SIP Spezifikation. Die einzelnen validierten Punkte werden in Gruppen wie z.B. Verzeichnisstruktur, Schema- und Prüfsummenvalidierung eingegliedert. Zuvor wird eine Formatvalidierung durchgeführt.

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.

Detailliertere Informationen zu den einzelnen Formaten und Validierungsschritte sind im Anhang ersichtlich.

2 Systemvoraussetzungen

Microsoft Windows 98 und neuer

Mindestens 128 MB RAM (512 MB oder mehr empfohlen)

Mindestens 20 GB Festplattenspeicher

Java Runtime Environment (JRE) Version 8 respektive 1.8

3 Offene Punkte / Feedback

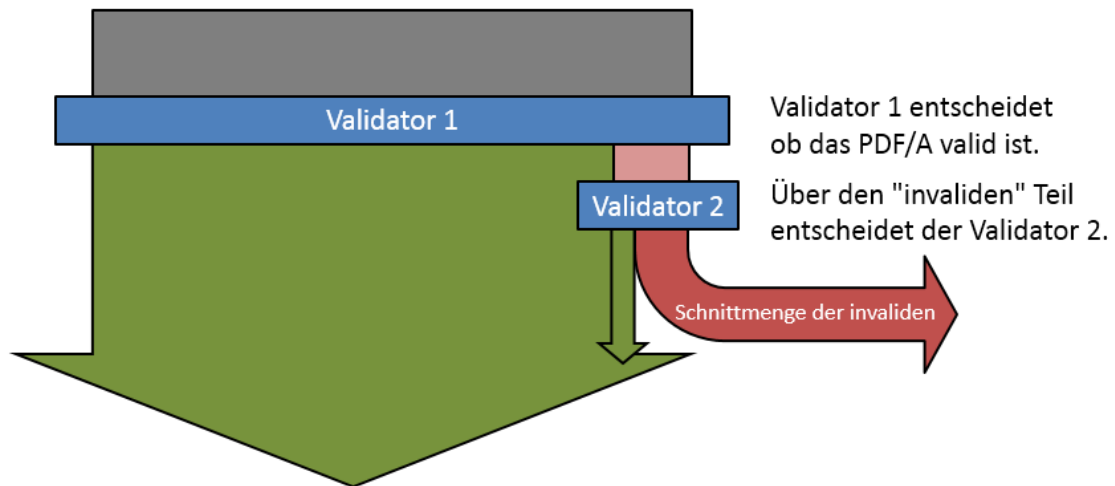
Die offenen Punkte, von Bugs respektive Fehlern bis hin zu Ergänzungswünschen und Fragen, sind auf der Entwicklungsplattform GitHub unter Issues ersichtlich (<https://github.com/KOST-CECO/KOST-Val/issues>) und können an kost-val@kost-ceco.ch gemeldet werden.

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

⁷ Die Spezifikation kann von der eCH-Website heruntergeladen werden:
<https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0160>.

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.⁸



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 (siehe 6.1.2) 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. Andererseits 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.

⁸ Die duale Validierung kann nur mit qualitativ hochstehenden PDF/A-Validatoren in diesem Sinne durchgeführt werden. Diese hohen Anforderungen erfüllen unter anderem die neuesten Versionen von 3-Heights™ PDF/A Validator von PDF-Tools und pdfaPilot von callas.

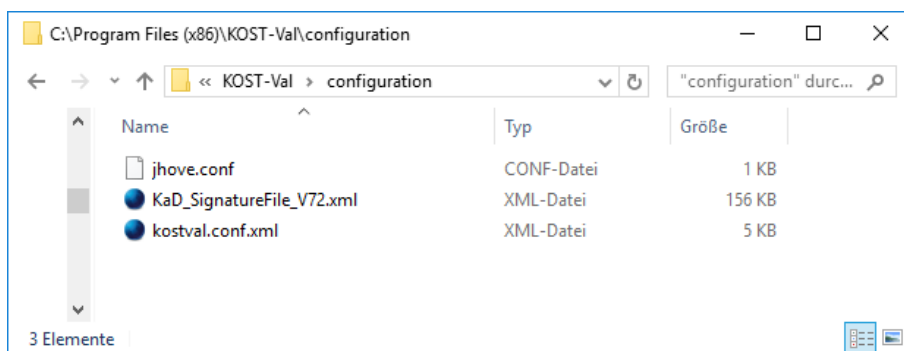
5 Installation von KOST-Val

- 1 KOST-Val herunterladen und in das gewünschte Verzeichnis entpacken.



Ab der KOST-Val Version 1.9.1 werden die dll-Dateien in der 64bit Version im KOST-Val-Verzeichnis beigelegt. Sollte nur eine 32bit JRE8 zur Verfügung stehen, müssen die dll-Dateien mit den Dateien aus dem Verzeichnis "resources\3-Heights_PDF-Validator_dll\Win32" ausgetauscht werden.

6 Konfiguration von KOST-Val



Im Ordner "configuration" sind die Dateien "jhove.conf" und "KaD_SignatureFile_V72.xml" abgelegt. Diese sind im Kapitel 6.2 beschrieben. Die Konfigurationsdatei "kostval.conf.xml" wird, falls nicht vorhanden, ins Verzeichnis "USERHOME/.kost-val/configuration" kopiert. Sämtliche Konfigurationen des KOST-Val können in der Konfigurationsdatei "USERHOME/.kost-val/configuration /kostval.conf.xml" vorgenommen werden. Für Modifikationen an XML-Dateien sollte ein XML-Editor oder Notepad ++⁹ verwendet werden.

6.1 Bestandteile der Konfigurationsdatei "kostval.conf.xml"

Die Konfigurationsdatei "kostval.conf.xml" ist in verschiedenen Teilen aufgebaut, welche im Kapitel 12.1 detailliert beschrieben werden.

Die ausgelieferte Konfiguration ermöglicht eine sofortige Validierung von PDF/A, TIFF, SIARD, JP2, JPEG und SIPs.

Nachfolgend werden die Bestandteile kurz beschrieben.

⁹ Die Portable Notepad ++ Applikation kann unter http://portableapps.com/de/apps/development/notepadpp_portable heruntergeladen werden.

6.1.1 Allgemeiner Teil

xml-Tag	Beschreibung: Initialwert
<showprogressonwork>	Anzeige, dass KOST-Val arbeitet: yes

6.1.2 PDF/A Teil

xml-Tag	Beschreibung: Initialwert
<pdftools>	Angabe, ob eine PDF/A-Validierung mit PDF Tools stattfinden soll: yes Wenn sowohl <callas> als auch <pdftools> auf yes eingestellt ist, wird eine duale Validierung durchgeführt (siehe auch Kapitel 4).
<detail>	Angabe, ob seitens PDF Tools auch detaillierte Fehler in Englisch ausgegeben werden sollen: yes
<callas>	Angabe, ob eine PDF/A-Validierung mit callas stattfinden soll: yes Wenn sowohl <callas> als auch <pdftools> auf yes eingestellt ist, wird eine duale Validierung durchgeführt (siehe auch Kapitel 4).
<nentry>	Angabe, ob seitens callas ein Fehler oder eine Warnung ausgegeben werden soll, wenn der N-Eintrag nicht übereinstimmt: W
<pdfa1>	Angabe, welche Konformität bei der Version 1 mindestens erreicht werden muss: 1B
<pdfa2>	Angabe, welche Konformität bei der Version 2 mindestens erreicht werden muss: 2U
<pdfafont>	Angabe, ob eine Schrift-Validierung (Durchsuch- und Extrahierbarkeit) stattfinden soll: no
<pdfaimage>	Angabe, ob eine Bild-Validierung (JPEG und JP2) stattfinden soll: no
<jbig2allowed>	Angabe, ob die JBIG2-Komprimierung ¹⁰ erlaubt ist: yes

¹⁰ Zur Problematik der JBIG2-Komprimierung siehe http://kost-ceco.ch/cms/index.php?jbig2-compression_de. Die Preservation-Planning-Expertengruppe PPEG der KOST empfiehlt, beim Erstellen von PDF-Dateien vorerst auf die Kompressionsart JBIG2 zu verzichten.

6.1.3 SIARD Teil

xml-Tag	Beschreibung: Initialwert
<siardvalidation>	Angabe, ob eine SIARD-Validierung stattfinden soll: yes

6.1.4 JP2 Teil

xml-Tag	Beschreibung: Initialwert
<jp2validation>	Angabe, ob eine JP2-Validierung stattfinden soll: yes

6.1.5 JPEG Teil

xml-Tag	Beschreibung: Initialwert
<jpegvalidation>	Angabe, ob eine JPEG-Validierung stattfinden soll: yes

6.1.6 TIFF Teil¹¹

xml-Tag	Beschreibung: Initialwert
<tiffvalidation>	Angabe, ob eine TIFF-Validierung stattfinden soll: yes
<allowedcompressionX>	Angabe, ob der Komprimierungsalgorithmus X erlaubt ist oder nicht. Nachfolgende sind im Initialzustand erlaubt: Uncompressed, CCITT 1D, T4/Group 3 Fax, T6/Group 4 Fax, LZW, PackBits
<allowedphotointerY>	Angabe, ob der Farbraum Y erlaubt ist oder nicht. Nachfolgende sind im Initialzustand erlaubt: WhitelsZero, BlackIsZero, RGB, RGB Palette
<allowedbitspersampleZ>	Angabe, ob Z Bits per Sample (pro Kanal) erlaubt sind oder nicht. Nachfolgende sind im Initialzustand erlaubt: 1, 4, 8, 16
<allowedmultipage>	Angabe, ob Multipage-TIFFs erlaubt sind oder nicht: 1
<allowedtiles>	Angabe, ob der Aufbau in Kacheln erlaubt ist oder nicht: 0
<allowedsize>	Angabe, ob Dateigrößen von 1000MB (~1GB) und grösser erlaubt sind oder nicht: 0

¹¹ Die Initialwerte basieren auf der KaD Preservation Planning Empfehlung (http://kost-ceco.ch/cms/index.php?preservation_de), welche unter anderem auf die Baseline-TIFF zurückgreift. Wo nötig wurden die Baseline-TIFFs präzisiert oder erweitert.

6.1.7 SIP Teil

xml-Tag	Beschreibung: Initialwert
<allowedlengthofpaths>	Erlaubte maximale Anzahl Zeichen in Pfadlängen: 179
<allowsipname>	Vorgaben zum Aufbau des SIP-Namens: SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_lw{3}
<allowedformats>	Auflistung der erlaubten Dateiformate, bestehend aus Unterelementen mit Detailinformationen. Nachfolgende sind im Initialzustand erlaubt: TXT, PDFA1, PDFA2, TIFF, JP2, JPEG, WAVE, MP3, MP4, MJ2, CSV, SIARD, WARC

6.1.8 Zu ignorierende Fehler

xml-Tag	Beschreibung: Initialwert
<ignore>	Hier können zu ignorierende Detailfehlermeldungen von PDF Tools eingegeben werden: The document does not conform to the requested standard. [PDF Tools: 0x83410612] The value of the key N is 4 but must be 3. [PDF Tools: 0x80410607]

6.2 "KaD_SignatureFile_V72.xml" und "jhove.conf"

Im Ordner "configuration" sind die Dateien "KaD_SignatureFile_V72.xml" und "jhove.conf" abgelegt, welche nicht angepasst werden müssen.

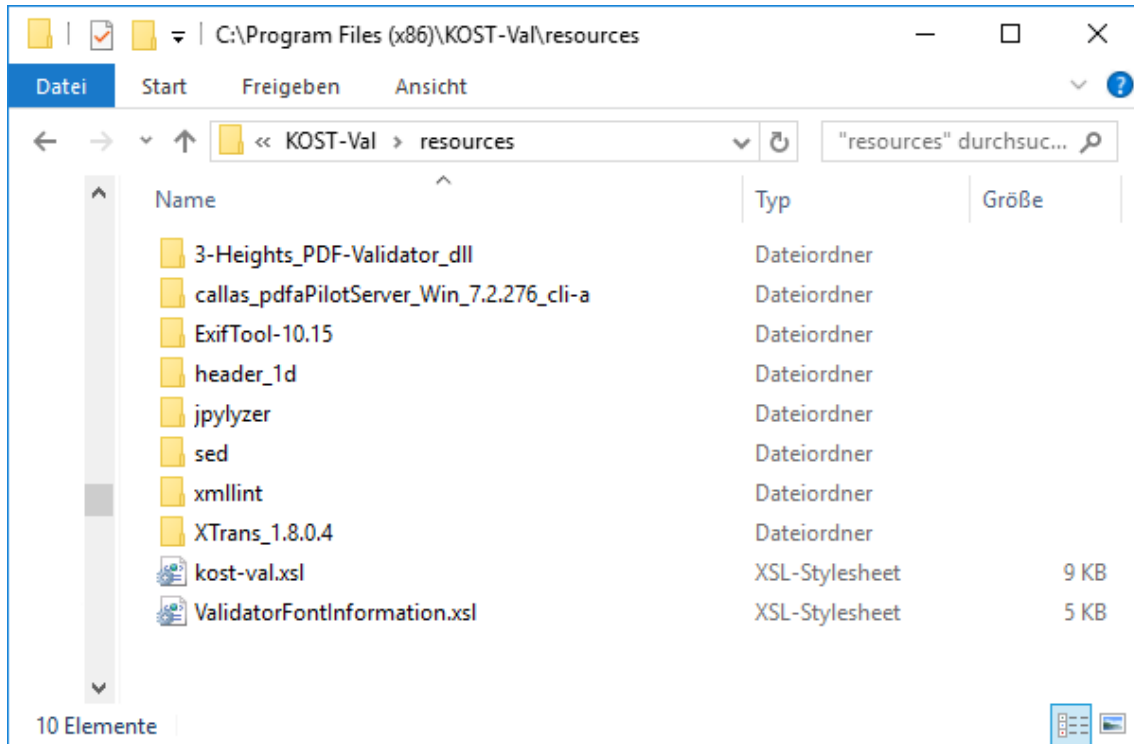
"KaD_SignatureFile_V72.xml" wird für die Formaterkennung benötigt. Es enthält den Stand von DROID per 28.08.2013 und wurde zudem durch die KOST im Bereich der Erkennung der archivtauglichen Formate¹² angepasst.

"jhove.conf" wird für die interne Jhove-Validierung benötigt.

¹² Anstelle des SignatureFile von DROID verwendet KOST-Val das dazu kompatible KaD-SignatureFile der KOST (siehe auch https://github.com/KOST-CECO/KaD_SignatureFile). Dieses umfasst nur die im KaD untersuchten Formate und ermöglicht deren Erkennung in der von der KOST empfohlenen Granularität. Gegenüber dem DROID-SignatureFile werden damit eine signifikante Effizienzsteigerung und eine höhere Benutzerfreundlichkeit erreicht.

7 Ressourcen von KOST-Val

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



8 Validierung starten



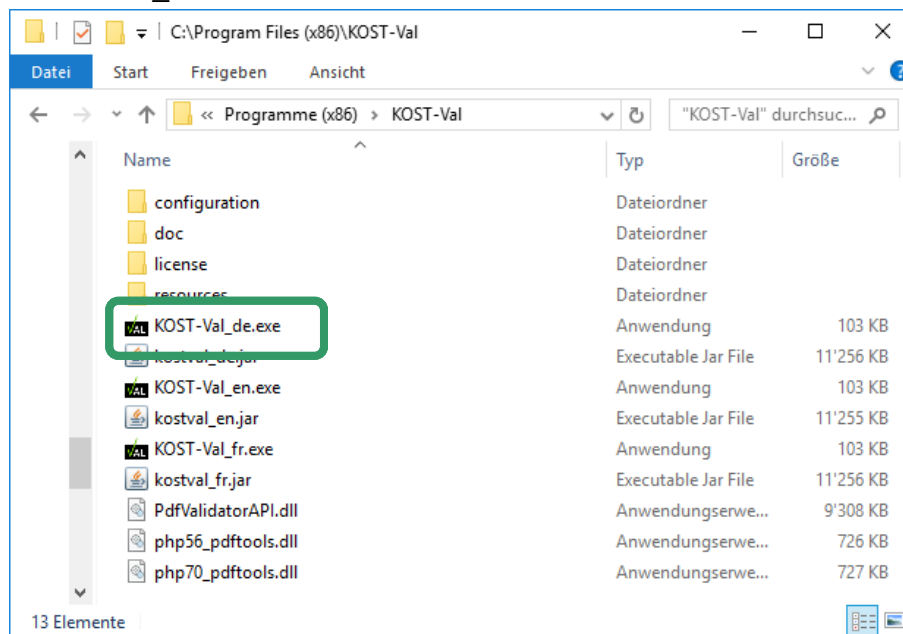
KOST-Val ist nicht Thread-sicher!

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 vorkommen.

8.1 Validierung starten mittels "KOST-Val_de.exe"-GUI

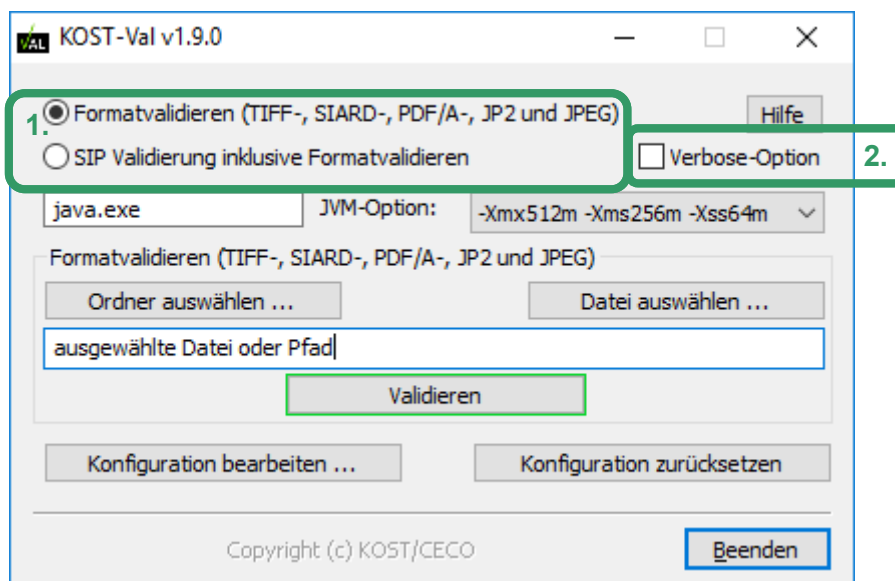
1

Starten von KOST-Val mittels Doppelklick auf "KOST-Val_de.exe" im Ordner "KOST-Val".

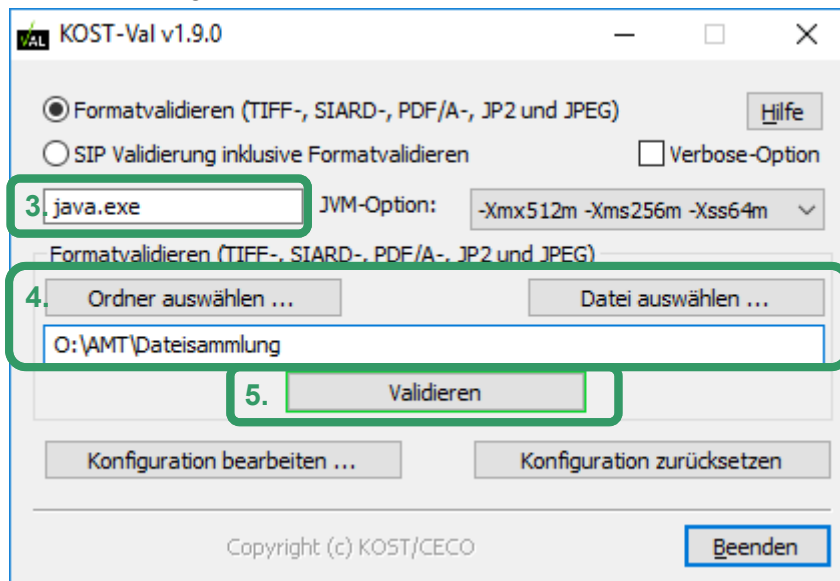


2

1. Auswählen, ob eine reine Formatvalidierung oder eine SIP-Validierung inkl. Formatvalidierung durchgeführt werden soll.
2. Auswählen, ob Reports von Jhove bzw. Jpylyzer behalten werden sollen oder nicht.



3. Den Pfad zu java.exe eingeben
4. Den Pfad zu der zu validierenden Datei oder zur Ordnerstruktur auswählen oder direkt eingeben.
5. Validierung starten

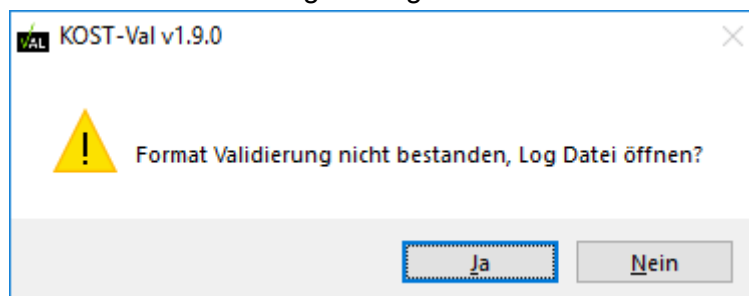


Hinweis:


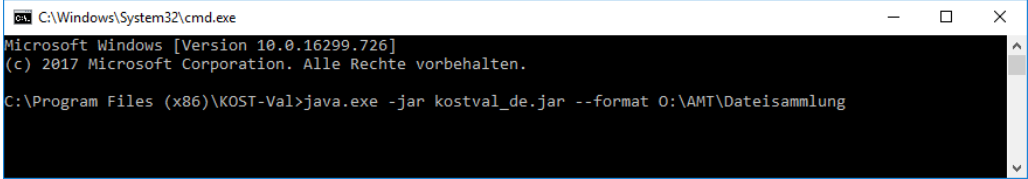
Bei Bedarf kann die Einstellung des virtuellen Java Memory angepasst werden. -Xmx sollte bei „Out of Memory“ und -Xss bei „Stack Overflow“ Fehlern angepasst werden.

Wenn gewünscht kann die Konfiguration temporär angepasst werden. Die temporäre Konfiguration wird beim Betätigen der Schaltflächen "Beenden" und "Konfiguration zurücksetzen" auf den ursprünglichen Stand zurückgesetzt.

- 3 Am Schluss der Validierung wird das Ergebnis angezeigt. Wenn gewünscht kann die KOST-Val-Log-Datei geöffnet werden.



8.2 Validierung manuell starten

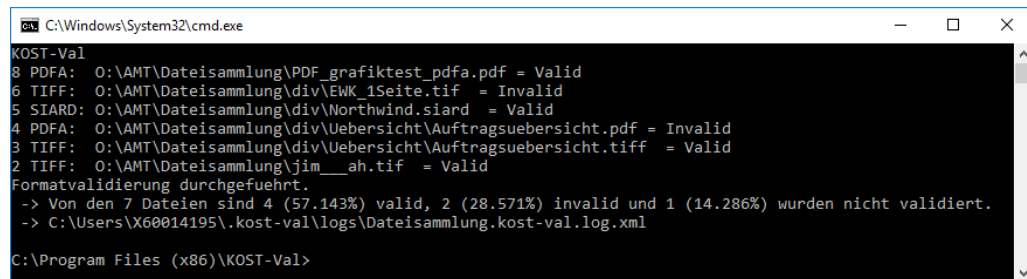
1	<p>Den zu validierenden Ordner oder Datei bereitstellen.</p> 
2	<p>Eingabeaufforderung öffnen und in das gewünschte Arbeitsverzeichnis wechseln (CD C:\Program Files (x86)\KOST-Val)¹³.</p>
3	<p>KOST-Val-Programmaufruf starten (die einzelnen Eingabebestandteile mit Leerzeichen trennen).</p>  <p>A java.exe¹⁴ -jar B kostval_de.jar C --format (oder --sip für die SIP-Validierung) D O:\AMT\Dateisammlung oder O:\AMT\Dateisammlung\jim__ah.tif E "nichts" oder -v (Verbose-Option)</p> <p>A = Java jre8 32bit Befehl (Pfad zu java.exe -jar), B = relationaler Pfad und die Datei kostval_de.jar, C = --format (oder --sip für die SIP-Validierung), D = relationaler Pfad und Name zum validierenden Ordner oder Datei, E = Verbose-Option, wenn die Originalreports behalten werden sollen.</p> <p><u>Anmerkungen:</u> Wenn ein Eingabebestandteil Leerzeichen enthält, muss dieser in Anführungs- und Schlusszeichen eingegeben werden. KOST-Val kann auch von einem beliebigen Ort aus aufgerufen werden. Dies bedingt jedoch die Eingabe von absoluten Pfaden.</p>

¹³ Das Laufwerk wird z.B. mit CD /D gewechselt.

¹⁴ Die Eingabe `java.exe -jar` ist nur möglich, wenn die gewünschte Java Runtime Environment (JRE) die Standardversion ist.

4

Die Datei wurde validiert, sobald "Valid" oder "Invalid" im cmd-Fenster erscheint. Der Ordner wurde validiert, sobald die Prompt (C:\Program Files (x86)\KOST-Val>) erscheint.



```
C:\Windows\System32\cmd.exe
KOST-Val
8 PDF: O:\AMT\Dateisammlung\PDF_grafiktest_pdfa.pdf = Valid
6 TIFF: O:\AMT\Dateisammlung\div\EWK_1Seite.tif = Invalid
5 SIARD: O:\AMT\Dateisammlung\div\Northwind.siard = Valid
4 PDF: O:\AMT\Dateisammlung\div\Uebersicht\Auftragsuebersicht.pdf = Invalid
3 TIFF: O:\AMT\Dateisammlung\div\Uebersicht\Auftragsuebersicht.tiff = Valid
2 TIFF: O:\AMT\Dateisammlung\jim__ah.tif = Valid
Formatvalidierung durchgefuehrt.
-> Von den 7 Dateien sind 4 (57.143%) valid, 2 (28.571%) invalid und 1 (14.286%) wurden nicht validiert.
-> C:\Users\X60014195\.kost-val\logs\Dateisammlung.kost-val.log.xml

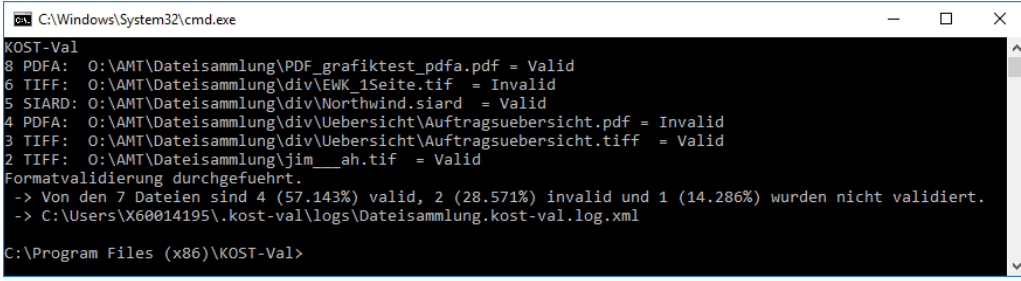
C:\Program Files (x86)\KOST-Val>
```

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

9 Validierung auswerten

1	Die Validierungsschritte sind im Kapitel 10 zusammengestellt.																																								
2	<p>Nach dem Start von KOST-Val werden die Validierungsart und die Datei angezeigt. Wenn der Validierung ohne Fehler abgeschlossen wurde, wird "Valid" ausgegeben.</p> <p>Sollte eine Validierung fehlerhaft sein, wird "Invalid" ausgegeben.</p>  <p>Wurde die Option "-v" (Verbose für "gesprächig") eingegeben, wird bei der Formatvalidierung zusätzlich auch der Originalreport ins Logverzeichnis geschrieben. Die Option "-v" ist nur für Formatspezialisten in seltenen Fällen interessant, da alle Fehlermeldungen 1:1 von KOST-Val übernommen werden.</p>																																								
3	<p>Die log-Datei führt zusätzliche Detailinformationen über die einzelnen invaliden Validierungsschritte auf, insbesondere der betroffene Validierungsschritt und der entsprechende Fehler.</p> <p>KOST-Val</p> <p>Format (PDF/A, TIFF, JP2, SIARD, JPEG):</p> <p>Von den 7 Dateien sind 4 (57.143%) valid, 2 (28.571%) invalid und 1 (14.286%) wurden nicht validiert. Folgende Dateien wurden nicht validiert (.db): - O:\AMT\Dateisammlung\Thumbs.db</p> <p>Invalid:</p> <p>Validierung: TIFF -> O:\AMT\Dateisammlung\div\EWK_1Seite.tif</p> <table border="1"> <tr> <td>B) Jhove</td> <td>Die Jhove-Validierung wurde nicht bestanden (Status: Well-Formed, but not valid).</td> </tr> <tr> <td>B) Jhove</td> <td>Jhove ErrorMessage: JPEGProc not defined for JPEG compression</td> </tr> <tr> <td>C) Komprimierung</td> <td>Compression: JPEG (old-style) ist nicht zulässig.</td> </tr> </table> <p>Validierung: PDF/A-2U -> O:\AMT\Dateisammlung\div\Uebersicht\Auftragsuebersicht.pdf</p> <table border="1"> <tr> <td>A) Allgemeines</td> <td>PDF/A-Eintrag nicht vorhanden [callas]</td> </tr> <tr> <td>C) Grafiken</td> <td>Geraeteabhngige Farbe verwendet aber kein PDF/A Output Intent vorhanden [callas]</td> </tr> <tr> <td>C) Grafiken</td> <td>Das Dokument enthlt geraetespezifische Farbraume. [PDF Tools: iCategory_3]</td> </tr> <tr> <td>C) Grafiken</td> <td>- A device-specific color space (DeviceGray) without an appropriate output intent is used. [PDF Tools: 0x03418614]</td> </tr> <tr> <td>D) Schriften</td> <td>Schrift nicht eingebettet (und Text-Darstellungsmodus nicht 3) [callas]</td> </tr> <tr> <td>D) Schriften</td> <td>CIDSet in Schrift-Untergruppe ist unvollstndig (Schrift enthlt Glyphen, die nicht gelistet sind) [callas]</td> </tr> <tr> <td>D) Schriften</td> <td>Das Dokument enthlt Schriften ohne eingebettete Font-Programme oder Codierungsinformation (CMAPs). [PDF Tools: iCategory_8]</td> </tr> <tr> <td>D) Schriften</td> <td>- The font ArialMT must be embedded. [PDF Tools: 0x00418704]</td> </tr> <tr> <td>D) Schriften</td> <td>- The font Arial-BoldMT must be embedded. [PDF Tools: 0x00418704]</td> </tr> <tr> <td>D) Schriften</td> <td>- The font Verdana must be embedded. [PDF Tools: 0x00418704]</td> </tr> <tr> <td>D) Schriften</td> <td>- The font Verdana-Bold must be embedded. [PDF Tools: 0x00418704]</td> </tr> <tr> <td>D) Schriften</td> <td>- The font Verdana-Italic must be embedded. [PDF Tools: 0x00418704]</td> </tr> <tr> <td>H) Metadaten</td> <td>Die Dokumentmetadaten sind nicht vorhanden, inkonsistent oder beschdigt. [PDF Tools: iCategory_16]</td> </tr> <tr> <td>H) Metadaten</td> <td>- The required XMP property 'pdfaid:part' is missing. [PDF Tools: 0x8341052B]</td> </tr> <tr> <td>H) Metadaten</td> <td>- The required XMP property 'pdfaid:conformance' is missing. [PDF Tools: 0x8341052B]</td> </tr> </table> <p>Valid:</p> <p>Validierung: PDF/A-1B -> O:\AMT\Dateisammlung\PDF_grafiktest_pdfa.pdf</p> <p>Validierung: SIARD-v1.0 -> O:\AMT\Dateisammlung\div\Northwind.siard</p> <table border="1"> <tr> <td>W) Warnung</td> <td>dataOwner enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.</td> </tr> <tr> <td>W) Warnung</td> <td>dataOriginTimespan enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.</td> </tr> </table> <p>Validierung: TIFF -> O:\AMT\Dateisammlung\div\Uebersicht\Auftragsuebersicht.tif</p> <p>Validierung: TIFF -> O:\AMT\Dateisammlung\jim_ah.tif</p>	B) Jhove	Die Jhove-Validierung wurde nicht bestanden (Status: Well-Formed, but not valid).	B) Jhove	Jhove ErrorMessage: JPEGProc not defined for JPEG compression	C) Komprimierung	Compression: JPEG (old-style) ist nicht zulässig.	A) Allgemeines	PDF/A-Eintrag nicht vorhanden [callas]	C) Grafiken	Geraeteabhngige Farbe verwendet aber kein PDF/A Output Intent vorhanden [callas]	C) Grafiken	Das Dokument enthlt geraetespezifische Farbraume. [PDF Tools: iCategory_3]	C) Grafiken	- A device-specific color space (DeviceGray) without an appropriate output intent is used. [PDF Tools: 0x03418614]	D) Schriften	Schrift nicht eingebettet (und Text-Darstellungsmodus nicht 3) [callas]	D) Schriften	CIDSet in Schrift-Untergruppe ist unvollstndig (Schrift enthlt Glyphen, die nicht gelistet sind) [callas]	D) Schriften	Das Dokument enthlt Schriften ohne eingebettete Font-Programme oder Codierungsinformation (CMAPs). [PDF Tools: iCategory_8]	D) Schriften	- The font ArialMT must be embedded. [PDF Tools: 0x00418704]	D) Schriften	- The font Arial-BoldMT must be embedded. [PDF Tools: 0x00418704]	D) Schriften	- The font Verdana must be embedded. [PDF Tools: 0x00418704]	D) Schriften	- The font Verdana-Bold must be embedded. [PDF Tools: 0x00418704]	D) Schriften	- The font Verdana-Italic must be embedded. [PDF Tools: 0x00418704]	H) Metadaten	Die Dokumentmetadaten sind nicht vorhanden, inkonsistent oder beschdigt. [PDF Tools: iCategory_16]	H) Metadaten	- The required XMP property 'pdfaid:part' is missing. [PDF Tools: 0x8341052B]	H) Metadaten	- The required XMP property 'pdfaid:conformance' is missing. [PDF Tools: 0x8341052B]	W) Warnung	dataOwner enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.	W) Warnung	dataOriginTimespan enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.
B) Jhove	Die Jhove-Validierung wurde nicht bestanden (Status: Well-Formed, but not valid).																																								
B) Jhove	Jhove ErrorMessage: JPEGProc not defined for JPEG compression																																								
C) Komprimierung	Compression: JPEG (old-style) ist nicht zulässig.																																								
A) Allgemeines	PDF/A-Eintrag nicht vorhanden [callas]																																								
C) Grafiken	Geraeteabhngige Farbe verwendet aber kein PDF/A Output Intent vorhanden [callas]																																								
C) Grafiken	Das Dokument enthlt geraetespezifische Farbraume. [PDF Tools: iCategory_3]																																								
C) Grafiken	- A device-specific color space (DeviceGray) without an appropriate output intent is used. [PDF Tools: 0x03418614]																																								
D) Schriften	Schrift nicht eingebettet (und Text-Darstellungsmodus nicht 3) [callas]																																								
D) Schriften	CIDSet in Schrift-Untergruppe ist unvollstndig (Schrift enthlt Glyphen, die nicht gelistet sind) [callas]																																								
D) Schriften	Das Dokument enthlt Schriften ohne eingebettete Font-Programme oder Codierungsinformation (CMAPs). [PDF Tools: iCategory_8]																																								
D) Schriften	- The font ArialMT must be embedded. [PDF Tools: 0x00418704]																																								
D) Schriften	- The font Arial-BoldMT must be embedded. [PDF Tools: 0x00418704]																																								
D) Schriften	- The font Verdana must be embedded. [PDF Tools: 0x00418704]																																								
D) Schriften	- The font Verdana-Bold must be embedded. [PDF Tools: 0x00418704]																																								
D) Schriften	- The font Verdana-Italic must be embedded. [PDF Tools: 0x00418704]																																								
H) Metadaten	Die Dokumentmetadaten sind nicht vorhanden, inkonsistent oder beschdigt. [PDF Tools: iCategory_16]																																								
H) Metadaten	- The required XMP property 'pdfaid:part' is missing. [PDF Tools: 0x8341052B]																																								
H) Metadaten	- The required XMP property 'pdfaid:conformance' is missing. [PDF Tools: 0x8341052B]																																								
W) Warnung	dataOwner enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.																																								
W) Warnung	dataOriginTimespan enthlt der Initialwert "(...)" und sollte korrekt ausgefllt werden.																																								

10 Beschreibung der Validierungsschritte

10.1 TIFF

Nachfolgend werden die einzelnen TIFF-Validierungsschritte detaillierter beschrieben.

A Erkennung

Wird die TIFF-Datei als TIFF erkannt?

B Jhove

Wurde die Jhove-Validierung bestanden?

C Komprimierung

Ist die verwendete Komprimierung gemäss der Konfiguration erlaubt?
Ist keine „Planar Configuration“ enthalten?

D Farbraum

Ist der verwendete Farbraum gemäss der Konfiguration erlaubt?

E BitsPerSample

Sind die verwendeten BitsPerSample gemäss der Konfiguration erlaubt?

F Multipage

Entspricht die Verwendung der Eigenschaft „Multipage“ der Konfiguration?

G Kacheln

Entspricht die Verwendung der Eigenschaft „Kacheln“ der Konfiguration?

H Grösse

Ist die Dateigrösse von der Konfiguration erlaubt?

10.2 SIARD

Nachfolgend werden die einzelnen Validierungsschritte detaillierter beschrieben. Diese Kriterien sind ein Auszug der Muss-Kriterien aus der SIARD-Spezifikation.

- A Lesbarkeit**
Kann die SIARD-Datei gelesen werden?
- B primäre Verzeichnisstruktur**
Besteht eine korrekte primäre Verzeichnisstruktur?
- C Header-Validierung**
Ist der header-Ordner valid?
- D Struktur-Validierung**
Stimmt die Struktur aus metadata.xml mit der Datei-Struktur von content überein?
- E Spalten-Validierung**
Wurden die Angaben aus metadata.xml korrekt in die tableZ.xsd-Dateien übertragen?
- F Zeilen-Validierung**
Wurden die Angaben aus metadata.xml korrekt in die tableZ.xsd-Dateien übertragen?
- G Tabellen-Validierung**
Sind die Spaltennamen innerhalb der Tabelle einmalig?
- H Content-Validierung**
Sind die XML-Dateien im content valid zu ihrer Schema-Definition (XSD-Dateien)?
- I SIARD-Erkennung**
Wird die SIARD-Datei als SIARD erkannt?
- J Zusätzliche Primärdateien**
Sind alle im content-Ordner enthaltenen Dateien oder Ordner in metadata.xml beschrieben?
- W Warnung**
Wurden <dataOwner> und <dataOriginTimespan> ausgefüllt und nicht auf (...) belassen?

10.3 PDF/A

Die einzelnen Meldungen im Originalreport von callas sowie die Detailfehlermeldungen von PDF Tools werden anhand einzelner Wörter der Fehlermeldung den Modulen zugeordnet. Die Fehlermeldungen (Message) werden 1:1 übernommen. Bei der Validierung mit PDF Tools werden die Kategorien den Modulen zugeordnet, übersetzt und der Text ausgegeben.

A Allgemeines

Sind die allgemeinen Anforderungen wie z.B. nicht korrupte PDF/A-Datei erfüllt?

B Struktur

Ist die PDF/A-Datei korrekt aufgebaut (z.B. BOF und EOF)?

C Grafiken

Sind die Grafiken korrekt in der PDF/A-Datei integriert (z.B. valider Farbraum)? Stimmt die Komponentenanzahl im N-Eintrag des PDF/A OutputIntent mit ICC-Profil überein?¹⁵

D Schriften

Sind die Schriften korrekt in der PDF/A-Datei eingebettet (z.B. sind alle verwendeten Zeichen eingebettet)?

E Transparenz

Sind die Anforderungen betreffend Transparenz an die PDF/A-Datei erfüllt?

F Annotationen & Interaktionen

Enthält die PDF/A-Datei nur erlaubte Annotationen (z.B. keine 3D-Annotation)? Enthält die PDF/A-Datei nur erlaubte Interaktionen?

G Aktionen

Enthält die PDF/A-Datei nur erlaubte Aktionen (z.B. kein JavaScript)?

H Metadaten

Sind die Metadaten korrekt in der PDF/A-Datei enthalten?

I Zugänglichkeit

Sind die Anforderungen an die Zugänglichkeit (Conformance Level A) erfüllt?

J Bild- & JBIG2-Validierung (konfigurierbar)

Sind die enthaltenen JPEG und JP2 valide? Wurde die JBIG2-Komprimierung verwendet?

K Schrift-Validierung (konfigurierbar)

Sind die enthaltenen Schriften durchsuch- und extrahierbar?

¹⁵ Dazu wird bei pdfaPilot von callas eine zusätzliche Prüfung durchgeführt, welche je nach Konfiguration (siehe 6.1.2) ggf. eine Warnung oder einen Fehler ausgibt. Bei 3-Heights™ PDF/A Validator von PDF Tools wird diese Prüfung standardmässig durchgeführt, jedoch in der iCategory_1 bemängelt, welches aufgrund der Zusammensetzung unter A) Allgemeines ausgegeben wird.

10.4 JP2

Die nicht bestandenen Tests von Jpylyzer werden den Modulen zugeordnet, und entsprechende Fehlermeldungen werden ausgegeben. Nachfolgend werden die einzelnen JP2-Validierungsschritte detaillierter beschrieben.

A Erkennung und Jpylyzer

Wird die JP2-Datei als JP2 erkannt? Wurde die Jpylyzer-Validierung bestanden?

B Metadaten-Validierung

Wurden die Metadaten korrekt erfasst?

C Bild-Validierung

Ist das Bild korrekt aufgebaut?

D Erweiterte Validierung

Wurden die zusätzlichen Jpylyzer-Tests bestanden?

10.5 JPEG

Die Fehlermeldungen aus BadPeggy werden den Modulen zugeordnet, und übersetzt in Deutsch ausgegeben (Original in Englisch). Nachfolgend werden die einzelnen JPEG-Validierungsschritte detaillierter beschrieben.

A Erkennung und BadPeggy

Wird die JPEG-Datei als JPEG erkannt? Wurde die BadPeggy-Validierung bestanden?

B Korrupte Daten

Gibt BadPeggy eine Fehlermeldung heraus, welche auf korrupte Daten hinweist?

C Ungültige Dateistruktur

Gibt BadPeggy eine Fehlermeldung heraus, welche auf eine ungültige Dateistruktur hinweist?

D Andere Probleme

Gibt BadPeggy eine Fehlermeldung aus, welche noch nicht den Modulen zugeordnet und übersetzt wurde?

10.6 SIP

Nachfolgend werden die einzelnen Validierungsschritte detaillierter beschrieben. Diese Kriterien sind ein Auszug der Muss-Kriterien aus der SIP-Spezifikation.

Modul 1: Paket- und XML-Konsistenz

1A Lesbarkeit

Kann das SIP (ZIP / ZIP64) fehlerfrei geöffnet werden?

1B primäre Verzeichnisstruktur

Besteht eine korrekte primäre Verzeichnisstruktur?

1C Verzeichnis- und Dateinamen

Entsprechen die Namen den Einschränkungen in der Spezifikation?

1D Schemavalidierung metadata.xml

Entspricht metadata.xml den Schemadateien in /header/xsd?

1E SIP-Typ ermitteln

Der SIP Typ wird ermittelt und angezeigt: GEVER oder FILE

1F Primärdateien im Verzeichnis

Sind Primärdateien im Verzeichnis /content vorhanden?¹⁶

Modul 2: Datei-Konsistenz

2A Fehlende Primärdateien

Sind alle referenzierten Dateien vorhanden?

2B Zusätzliche Primärdateien

Sind im SIP keine zusätzlichen Primärdateien vorhanden?

2C Prüfsummen-Validierung

Stimmen die Prüfsummen der Dateien mit Prüfsumme überein?

2D Verzeichnung GEVER-Primärdateien

Sind alle referenzierten Dateien auch im Ordnungssystem verzeichnet?

Modul 3: Dateiformat- und Datums-Konsistenz

3A Formaterkennung

Sind die erkannten Formate erlaubt?

3B Zusätzliche Formate

Alle Dateien in nicht erlaubten Formaten mit entsprechenden Formatangaben auflisten.

3C Formatvalidierung

Sind die Formate valid?

3D Zeitraum-Validierung

Stimmen die Zeitangaben in (metadata.xml)/ablieferung überein?

¹⁶ Primärdateien können in einem GEVER-SIP fehlen, das nur zur Archivierung einer Ordnerstruktur dient, aber nicht in einem FILE-SIP.

11 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 Foundation (http://www.apache.org/).
----------------	---

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	4.12.15	Siehe Kapitel 11.1
Apache Commons http://commons.apache.org/ - commons-logging-1.2.jar - commons-lang-3.7.jar - commons-logging-1.2.jar	1.2 3.7 1.2	Apache License 2.0
Apache log4j http://logging.apache.org/log4j/	1.2.12	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
DROID http://digital-preservation.github.io/droid/	5.0.3	3c BSD- License
iText http://itextpdf.com/	5.5.5	AGPL v3
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 4.12 http://www.junit.org/	4.12	CPL v1.0
Spring Framework API http://static.springsource.org/spring/docs/5.0.x/api/	5.0.8	Apache License 2.0
zip64 http://sourceforge.net/projects/zip64file/	1.02	GPL v2+ License

KOST-Val stützt sich auf folgende unveränderte Komponenten anderer Hersteller, welche mit KOST-Val abgegeben werden:

Drittprogramm / -Komponente	Version	Lizenz
ExifTool http://www.sno.phy.queensu.ca/~phil/exiftool/	10.15	PERL respektive GPL v3.0 License
Jpylyzer http://jpylyzer.openpreservation.org/	1.17.0	LGPL v3.0 License
NSIS v2.46 http://nsis.sourceforge.net/Main_Page	2.46	zlib/libpng License
pdfaPilot CLI https://www.callassoftware.com	7.2	Siehe Kapitel 11.2
GNU sed https://www.gnu.org/software/sed	4.4	GPL v3+ License
XML.nsh http://nsis.sourceforge.net/XML_plugin	2.0	zlib/libpng License
XmlLint http://xmlsoft.org/xmlLint.html	20630	MIT License
XTrans http://sourceforge.net/projects/xtrans/	1.8.0.4	GPL v2 License

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

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

11.2 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.

12 Anhang

12.1 Beschreibung der Konfigurationsdatei "kostval.conf.xml"

Die Konfigurationsdatei "kostval.conf.xml" ist aus verschiedenen Teilen aufgebaut, welche nachfolgend detailliert beschrieben werden.

12.1.1 Allgemeiner Teil

<showprogressonwork>

Angabe, ob angezeigt (yes) werden soll oder nicht (no), dass KOST-Val arbeitet. Der Initialwert ist **yes**.

12.1.2 PDF/A-Teil

<pdftools>

Angabe, ob eine PDF/A-Validierung mit PDF Tools stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

Wenn sowohl <callas> als auch <pdftools> auf yes eingestellt ist, wird eine duale Validierung durchgeführt (siehe auch Kapitel 4).

<detail>

Angabe, ob seitens PDF Tools auch detaillierte Fehlermeldungen in Englisch ausgegeben werden sollen (detail) oder nicht (no). Der Initialwert ist **detail**.

<callas>

Angabe, ob eine PDF/A-Validierung mit callas stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

Wenn sowohl <callas> als auch <pdftools> auf yes eingestellt ist, wird eine duale Validierung durchgeführt (siehe auch Kapitel 4).

<nentry>

Angabe, ob seitens callas ein Fehler (E) oder eine Warnung (W) ausgegeben werden soll, wenn der N-Eintrag nicht übereinstimmt. Der Initialwert ist **W**.

<pdfa1>

Angabe, welche Konformität in der Version 1 mindestens erreicht werden muss. Dabei kann gewählt werden zwischen *Accessible* (1A), *Basic* (1B) oder nicht erlaubt (no). Der Initialwert ist **1B**.

<pdfa2>

Angabe, welche Konformität in der Version 2 mindestens erreicht werden muss. Dabei kann gewählt werden zwischen *Accessible* (2A), *Basic* (2B), *Unicode* (2U) oder nicht erlaubt (no). Der Initialwert ist **2U**.

<pdfaimage>

Angabe, ob eine Bildvalidierung stattfinden soll (yes) oder nicht (no). Bei der Bildvalidierung werden die enthaltenen JPEG und JP2 extrahiert und validiert. Der Initialwert ist **no**.

<pdfafont>

Angabe, ob eine Schriftvalidierung stattfinden soll (tolerant, warning oder strict) oder nicht (no). Bei der Schriftvalidierung wird untersucht, ob die enthaltenen Zeichen durchsuch- und extrahierbar sind. Der Initialwert ist **no**; empfohlener Wert für die Schriftvalidierung ist **warning**.

- no keine Schriftvalidierung
- tolerant Schriftvalidierung; Symbole und undefinierte Zeichen werden ignoriert, sofern mindestens 80% aller Zeichen korrekt eingebettet sind
- warning Schriftvalidierung, wobei Symbole und undefinierte Zeichen eine Warnung ergeben, sofern mindestens 90% aller Zeichen korrekt eingebettet sind.
- strict Schriftvalidierung; Symbole, undefinierte und unbekannte Zeichen sind Fehler

<jbig2allowed>

Angabe, ob die JBIG2-Komprimierung erlaubt ist (yes) oder nicht (no). Der Initialwert ist **yes**.

12.1.3 SIARD-Teil

<siardvalidation>

Angabe, ob eine SIARD-Validierung stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

12.1.4 JP2-Teil

<jp2validation>

Angabe, ob eine JP2-Validierung stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

12.1.5 JPEG-Teil

<jpegvalidation>

Angabe, ob eine JPEG-Validierung stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

12.1.6 TIFF-Teil

Die Initialwerte basieren auf der KaD Preservation Planning Empfehlung (http://kost-ceco.ch/cms/index.php?preservation_de), welche unter anderem auf Baseline TIFF zurückgreift. Wo nötig, wurde Baseline TIFF präzisiert oder erweitert.

<tiffvalidation>

Angabe, ob eine TIFF-Validierung stattfinden soll (yes) oder nicht (no). Der Initialwert ist **yes**.

<allowedcompressionX>

Angabe, ob der Komprimierungsalgorithmus erlaubt ist (Beschreibung) oder nicht (0). Die initial erlaubten Komprimierungsalgorithmen sind **Uncompressed**, **CCITT 1D**, **T4/Group 3 Fax**, **T6/Group 4 Fax**, **LZW**, **PackBits**.

X	Beschreibung	Initialwert	Herkunft
1	Uncompressed	Uncompressed	Baseline-TIFF
2	CCITT 1D	CCITT 1D	Baseline-TIFF
3	T4/Group 3 Fax	T4/Group 3 Fax	Baseline-TIFF
4	T6/Group 4 Fax	T6/Group 4 Fax	Baseline-TIFF
5	LZW	LZW	KaD-PP Erweiterung
7	JPEG	0	TIFF v6 TN2
8	Deflate	0	TIFF v6
32773	PackBits	PackBits	Baseline-TIFF

<allowedphotointerY>

Angabe, ob der Farbraum erlaubt ist (Beschreibung) oder nicht (0). Die initial erlaubten Farbräume sind **WhitelsZero**, **BlackIsZero**, **RGB**, **RGB Palette**.

Y	Beschreibung	Initialwert	Herkunft
0	WhitelsZero	WhitelsZero	Baseline-TIFF
1	BlackIsZero	BlackIsZero	Baseline-TIFF
2	RGB	RGB	Baseline-TIFF
3	RGB Palette	RGB Palette	Baseline-TIFF
4	transparency mask	0	TIFF v6
5	CMYK	0	TIFF v6
6	YCbCr	0	TIFF v6
7	CIE L*a*b*	0	TIFF v6

<allowedbitspersampleZ>

Angabe, ob die Bits per Sample (pro Kanal) erlaubt sind (Z) oder nicht (0). Die initial erlaubten Bits per Sample sind **1**, **4**, **8**, **16**.

Z	Beschreibung	Initialwert	Herkunft
1	1 Bits per Sample (pro Kanal)	1	Baseline-TIFF
2	2 Bits per Sample (pro Kanal)	0	Baseline-TIFF
4	4 Bits per Sample (pro Kanal)	4	Baseline-TIFF
8	8 Bits per Sample (pro Kanal)	8	Baseline-TIFF
16	16 Bits per Sample (pro Kanal)	16	KaD-PP Erweiterung
32	32 Bits per Sample (pro Kanal)	0	Baseline-TIFF
64	64 Bits per Sample (pro Kanal)	0	Baseline-TIFF

<allowedmultipage>

Angabe, ob Multipage-TIFFs erlaubt sind (1) oder nicht (0). Der Initialwert ist **1** und entspricht der KaD Preservation Planning Empfehlung TIFF.

<allowedtiles>

Angabe, ob der Aufbau in Kacheln erlaubt ist (1) oder nicht (0). Der Initialwert ist **0** und entspricht den Vorgaben der Baseline-TIFFs.

<allowedsize>

Angabe, ob Dateigrößen von 1000MB (~1GB) und grösser erlaubt sind (1) oder nicht (0). Der Initialwert ist **0** und entspricht der KaD Preservation Planning Empfehlung TIFF.

12.1.7 SIP Teil

<allowedlengthofpaths>

Erlaubte maximale Anzahl Zeichen in Pfadlängen. Der Initialwert ist **179** und entspricht der Einschränkung in der SIP-Spezifikation: weniger als 180 Zeichen.

<allowsipname>

Vorgaben zum Aufbau des SIP-Namens. Der Initialwert (Regex-Ausdruck) ist **SIP_[1-2][0-9]{3}[0-1][0-9][0-3][0-9]_\\w{3}** und entspricht der Einschränkung i in der SIP-Spezifikation: SIP_[Ablieferungsdatum]_[Name der Abliefernden Stelle](_[Referenz]), wobei das Datum in der Form JJJJMMTT angegeben werden muss und die Referenz optional verwendet werden kann.

<allowedformats>

Auflistung der erlaubten Dateiformate bestehend aus Unterelementen mit Detailinformationen. Die Initialwerte sind aus der nachfolgenden Tabelle ersichtlich. Die Liste ist bewusst ausführlich gehalten, so dass die nicht erlaubten Formate einfach entfernt werden können und möglichst wenig nachgetragen werden muss. Angaben betreffend der Extension und der PUID sind der KaD_SignatureFile_V72.xml zu entnehmen.

<puid> </puid>	<extension> </extension>	<comment> </comment>
TXT	txt	Plain Text file
PDFA1	pdf	PDF/A-1
PDFA2	pdf	PDF/A-2
TIFF	tif, tiff	Tagged Image File Format
JP2	jp2, jpg2, j2c	JPEG2000
JPEG	jpeg, jpg, jpe	JPEG
SVG	svg	Scalable Vector Graphics
WAVE	wav, wave	Waveform Audio
MP3	mp3	MPEG 1/2 Audio Layer 3
MP4	mp4, m4v, m4a, f4v, f4a	MPEG-4 Media File
MJ2	mj2, mjp2	MJ2 (Motion JPEG 2000)
MPEG2	mpeg, mpg, m2v	MPEG-2 Video Format
PRORES	mov	Apple ProRes
CSV	csv	Comma Separated Values
SIARD	siard	SIARD
XML	xml	Ext. Markup Language
XSD	xsd	XML Schema Definition
XLS	xls, xlw	Microsoft Excel (xls)
WARC	warc	WARC
ARC	arc	Internet Archive
HTML	htm, html	Hypertext Markup Language

12.1.8 Zu ignorierende Fehler

<ignore>

Hier können zu ignorierende Detailfehlermeldungen von PDF Tools eingegeben werden. Der Initialwert ist: **The document does not conform to the requested standard. [PDF Tools: 0x83410612] The value of the key N is 4 but must be 3. [PDF Tools: 0x80410607]**

12.2 Programmaufbau

KOST-Val wurde nach folgenden Anforderungen aufgebaut:

Funktionale Anforderungen:

TIFF-Validierung: KOST-Val liest ein TIFF und validiert mit Hilfe von Jhove die folgenden Punkte:

Validierungsschritt	Bezeichnung
A (Abbruch wenn Fehler)	Erkennung
B	Jhove
C	Komprimierung
D	Farbraum
E	BitsPerSample
F	Multipage
G	Kacheln
H	Grösse

SIARD-Validierung: KOST-Val liest ein SIARD und validiert die folgenden Punkte:

Validierungsschritt	Bezeichnung
A (Abbruch wenn Fehler)	Lesbarkeit
B (Abbruch wenn Fehler)	Primäre Verzeichnisstruktur
C (Abbruch wenn Fehler)	Header-Validierung
D (Abbruch wenn Fehler)	Struktur-Validierung
E	Spalten-Validierung
F	Zeilen-Validierung
G	Tabellen-Validierung
H	Content-Validierung
I	SIARD-Erkennung
J	Zusätzliche Primärdateien
W	Warnung

PDF/A-Validierung: KOST-Val liest ein PDF und validiert mit Hilfe des 3-Heights™ PDF/A Validator von PDF-Tools oder auch des pdfaPilot von callas die folgenden Punkte:

Validierungsschritt	Bezeichnung
A (Abbruch wenn Fehler)	Allgemeines
B	Struktur
C	Grafiken
D	Schriften
E	Transparenz
F	Annotationen
G	Aktionen & Interaktionen
H	Metadaten
I	Zugänglichkeit
J (konfigurierbar)	Bild- & JBIG2-Validierung
K (konfigurierbar)	Schrift-Validierung

JP2-Validierung: KOST-Val liest ein JP2 und validiert mit Hilfe von Jpylyzer die folgenden Punkte:

Validierungsschritt	Bezeichnung
A (Abbruch möglich wenn Fehler)	Erkennung und Jpylyzer
B	Metadaten-Validierung
C	Bild-Validierung
D	Erweiterte Validierung

JPEG-Validierung: KOST-Val liest ein JPEG und validiert mit Hilfe von BadPeggy die folgenden Punkte:

Validierungsschritt	Bezeichnung
A (Abbruch möglich wenn Fehler)	Erkennung und BadPeggy
B	Korrupte Daten
C	Ungültige Dateistruktur
D	Andere Probleme

SIP-Validierung: KOST-Val liest ein SIP und validiert die folgenden Punkte aus der SIP Spezifikation:

Validierungsschritt	Bezeichnung (Stepname)
1a (Abbruch wenn Fehler)	Lesbarkeit
1b (Abbruch wenn Fehler)	primäre Verzeichnisstruktur
1c (Abbruch wenn Fehler)	Verzeichnis- und Dateinamen
1d (Abbruch wenn Fehler)	Schemavalidierung metadata.xml
1e	SIP-Typ ermitteln
1f	Primärdateien im Verzeichnis
2a	Fehlende Primärdateien
2b	Zusätzliche Primärdateien
2c	Prüfsummen-Validierung
2d	Verzeichnung GEVER-Primärdateien
3a	Formaterkennung
3b	Zusätzliche Formate
3c	Formatvalidierung
3d	Zeitraum-Validierung

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.

12.3 Funktionsprinzip Formatvalidierung

