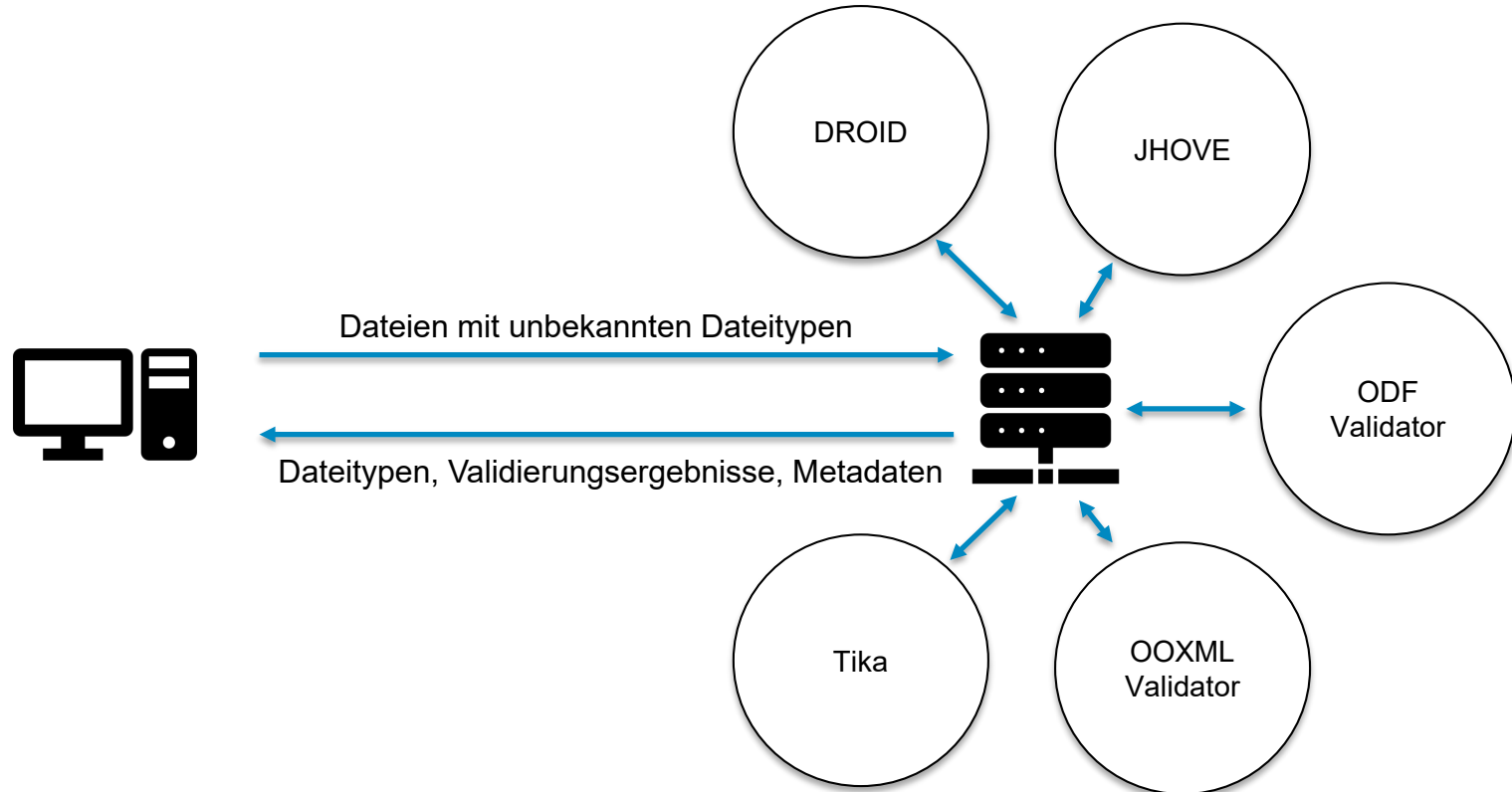


Borg

Open Source-Programm des Landesarchivs Thüringen
zur einfacheren Einbindung und Kombination beliebiger
Formaterkennungs- und Validierungswerkzeuge

Tony Grochow, Landesarchiv Thüringen

- Konfigurierbare Schnittstelle zur einfachen Einbindung und Kombination beliebiger Formaterkennungs- und Validierungswerkzeuge
- eingebundene Werkzeuge werden in Docker-Containern ausgeführt
 - geringe Abhängigkeit von Systemanforderungen der Werkzeuge
- Werkzeugergebnisse werden nach konfigurierbaren Regeln zu einem Gesamtergebnis zusammengefasst
- Export des Gesamtergebnisses inkl. der Einzelergebnisse der Werkzeuge möglich
- Integration in andere Anwendungen möglich (Web-API)













Integration umgesetzt:

- DROID
- Tika
- JHOVE
- veraPDF
- ODF Validator
- OOXML Validator

Integration geplant:

- Magika
- Checkit tiff
- Jpylyzer
- ...

Pfad ↑	Dateiname	Dateigröße	Upload	Formatverifikation
Testdaten_Borg/Bilddateien/	09_E.jpg	915 B	✓	wird durchgeführt 
Testdaten_Borg/Bilddateien/	12_æ.tif	522 B	✓	wird durchgeführt 
Testdaten_Borg/Bilddateien/	kakadu61.jp2	653,7 KB	✓	wird durchgeführt 
Testdaten_Borg/CSV/	wetterdaten_Wetterstation.csv	3,11 KB	✓	wird durchgeführt 
Testdaten_Borg/Dokumente/	header-intakt_random-bitstream_body.pdf	102,32 KB	✓	wird durchgeführt 
Testdaten_Borg/Dokumente/	Information_Anfragen_und_Benutzer-20.10.23.pdf	179,67 KB	✓	wird durchgeführt 
Testdaten_Borg/Dokumente/	invalid_isartor-6-3-5-t02-fail-a.pdf	19,09 KB	wartet auf Serverkapazität 	
Testdaten_Borg/Dokumente/	test_pdf_a1-b.pdf	64,02 KB	wartet auf Serverkapazität 	
Testdaten_Borg/Dokumente/	test_pdf_a2-b.pdf	99,33 KB	wartet auf Serverkapazität 	
Testdaten_Borg/Web/	html-4.01strict.html	809 B	wartet auf Serverkapazität 	

[📄 Einzeldatei hinzufügen](#)[📁 Ordner hinzufügen](#)


Einträge pro Seite

10

1 - 10 von 11

[<<](#) [<](#) [>](#) [>>](#)

Pfad	Dateiname	Dateigröße	PUID	MIME-Type	Formatversion	Status
Testdaten_Borg/Bilddateien/	09_E.jpg	915 B	fmt/43	image/jpeg	JPEG 1.01	✓
Testdaten_Borg/Bilddateien/	kakadu61.jp2	653,7 KB	x-fmt/392	image/jp2		✓
Testdaten_Borg/Bilddateien/	12_æ.tif	522 B	fmt/353	image/tiff	TIFF 6.0	✓
Testdaten_Borg/Dokumente/	header-intakt_random-bitstream_body.pdf	102,32 KB	fmt/18			⚠️
Testdaten_Borg/Dokumente/	Information_Anfragen_und_Benutzer-20.10.23.pdf	179,67 KB	fmt/276	application/pdf	1.7	✓
Testdaten_Borg/Web/	html-4.01strict.html	809 B	fmt/100	text/html	HTML 4.01	✓
Testdaten_Borg/Web/	html-4.01strict_invalide.html	808 B	fmt/100	text/html	HTML 4.01	✗
Testdaten_Borg/Dokumente/	invalid_isartor-6-3-5-102-fail-a.pdf	19,09 KB	fmt/354	application/pdf	PDF/A-1b	✗
Testdaten_Borg/Dokumente/	test_pdf_a1-b.pdf	64,02 KB	fmt/354	application/pdf	PDF/A-1b	✓
Testdaten_Borg/Dokumente/	test_pdf_a2-b.pdf	99,33 KB	fmt/20	application/pdf	PDF/A-2b	✓

 Auswertung leeren Ergebnisse exportieren

Einträge pro Seite

15 ▾

16 - 25 von 25



Formatverifikation

✓ Der Datei ist eine valide Datei ihres Typs.

Werkzeug	PUID	MIME-Type	Formatversion	Valide
Gesamtergebnis	fmt/20 (90 %)	application/pdf (90 %)	PDF/A-2b (90 %)	✓ (77 %)
DROID	fmt/20 (90 %)			
Tika		application/pdf (90 %)	PDF/A-2b (90 %)	
JHOVE (PDF-Module)			PDF 1.6 (0 %)	✓ (0 %)
veraPDF (PDF/A-2b-Profile)				✓ (100 %)
veraPDF (PDF/UA-Profile)				✗ (30 %)

Formatverifikation

✘ Der Datei ist invalide.

Die Ursache kann eine Erstellung mit ungeeigneten Werkzeugen oder eine Beschädigung der Datei sein.

Dies bedeutet nicht automatisch, dass die Datei unbrauchbar ist.

Werkzeug	PUID	MIME-Type	Formatversion	Valide
Gesamtergebnis	fmt/354 (90 %)	application/pdf (90 %)	PDF/A-1b (90 %)	✘ (100 %)
DROID	fmt/354 (90 %)			
Tika		application/pdf (90 %)	PDF/A-1b (90 %)	
veraPDF (PDF/A-1b-Profile)				✘ (100 %)
veraPDF (PDF/UA-Profile)				✘ (30 %)

Tika (2.9.0)

Extrahierte Eigenschaften

Eigenschaft	Wert
Zeichenkodierung	ISO-8859-1
MIME-Type	text/html

Werkzeug-Ausgabe

Werkzeug-Ausgabe

```
{
  "X-TIKA:Parsed-By": [
    "org.apache.tika.parser.DefaultParser",
    "org.apache.tika.parser.html.HtmlParser"
  ],
  "dc:title": "Untitled Document",
  "Content-Encoding": "ISO-8859-1",
  "Content-Type-Hint": "text/html; charset=iso-8859-1",
  "resourceName": "b9fa892e-463c-48db-95fd-f423545ade68_html-4.01strict.html",
  "Content-Length": "809",
  "Content-Type": "text/html; charset=ISO-8859-1"
}
```

Kostenlose Nachnutzung durch andere Archive/Einrichtungen möglich

- Open Source-Lizenz
- Download ab 13. März 2024 unter <https://github.com/Landesarchiv-Thueringen>
- Textbausteine aus dem gemäß DSGVO notwendigen Verzeichnis der Verarbeitungstätigkeiten (VVT) können vom Landesarchiv zur Verfügung gestellt werden.
- Wir freuen uns über Anwender, Mitentwickler, Hinweise, Kritik, Umsetzungswünsche und Fachdiskussionen zum Programm!

Ansprechpartner für Borg: Tony Grochow (tony.grochow@la.thueringen.de)
Christine Träger (christine.traeger@la.thueringen.de)