Freistaat Thüringen Landesarchiv

Echtheitsnachweis to go.

Werkzeug zur Integritätswahrung elektronischer Reproduktionen

24. AUdS – 23.03.2021

Tony Grochow / Konrad Meckel

Landesarchiv Thüringen & Bauhaus-Universität Weimar

Langjährige Kooperation zwischen

- Projekt Digitales Magazin des Freistaats Thüringen zur Vorbereitung und Einführung der elektronischen Archivierung für alle Landesbehörden
- Bauhaus Universität, Lehrstuhl für Mediensicherheit

im Rahmen der Kooperation:

- Auseinandersetzung mit unterschiedlichen Aspekten und Herausforderungen der digitalen Archivierung
- u.a. Erstellung von Bedrohungsszenarien für Archivgut und dessen Benutzung
- u.a. mehrere Vorträge im Arbeitskreis:
  - Prof. Stephan Lucks: Kryptographie und Fehlertoleranz für Digitale Magazine; 2013 Dresden
  - Prof. Stephan Lucks: Digitale Magazine ohne eigenen Speicher.
    Wie man die Integrität “fremdgespeicherter” Archivalien sicherstellen kann; 2014 Weimar
  - Veronika Krauss: Authentizität digitaler Objekte unter Formattransformationen; 2016 Potsdam

Szenario

- Archivbenutzer erhält eine Reproduktion eines Nutzungspakets aus dem Digitalen Archiv („DIP2Go“)
  → Kopie z.B. auf USB-Stick oder über Datenaustauschplattform
- Archivbenutzer bearbeitet sein Thema über einen längeren Zeitraum
- Speicherung erfolgt i.d.R. in unsichere Umgebung
  → Veränderung oder Manipulation einfach möglich
- Sind vom Archiv erhaltenen Daten noch vollständig und unverändert?

→ Tool zur Integritätsprüfung ermöglicht Kontrolle

Tool

- Java-Programm (Java, Java FX)
    - Erstentwicklung: Christof Bräutigam, Student der Bauhaus-Universität Weimar
    - Aktualisierung: Tony Grochow

- Theoretische Grundlage:

  Vortrag von Prof. Lucks 2014: Absicherung über Hashbäume
    - jedoch auf den Anwendungsfall „DIP2Go“ beschränkt
    - Absicherung mit geringem zusätzlichen Speicheraufwand
    - Erfordert kein Schlüsselmanagement, keine Neusignierung und kein Geheimnis

Hashfunktion?

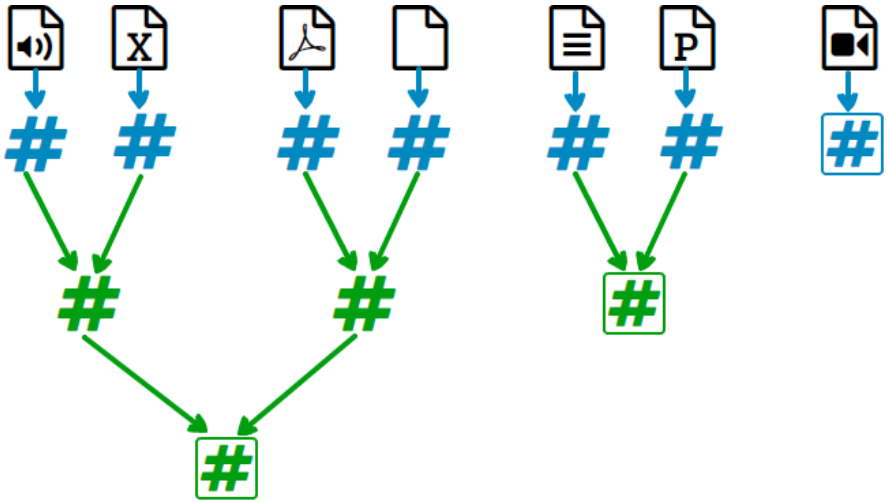
- über kryptographische Hashfunktion berechneter (kleiner) Fingerabdruck zu einer Datei oder Zeichenfolge

- Beispiel:
  ursprüngliche Textdatei, Inhalt: 1.000x Buchstabe „a“:
  Hashwert: `41EDECE42D63E8D9BF515A9BA6932E1C20CBC9F5A5D134645ADB5DB1B9737EA3`

  geänderte Textdatei, Inhalt: 1x Buchstabe „b“, dann 999x Buchstabe „a“:
  Hashwert: `EB7F72A09B36323AF46C121578EE51F161AA40C76DB8BD942420233A7A61DDC6`

→ über Hashfunktionen lassen sich einfach Veränderungen an Daten und Dateien feststellen

Hashbäume?



1. Zu jeder Datei wird ein Hashwert berechnet (blau). wichtig: die Reihenfolge muss immer gleich sein
2. aus je 2 benachbarten Hashwerten wird ein weiterer Hashwert berechnet. Es entsteht ein binärer Baum (grün).
3. Die Äste eines Baumes sind ausgeglichen. Ansonsten entsteht ein weiterer Baum.

zur Überprüfung genügen die Wurzelwerte der vorhandenen Bäume (eingerahmt: #).

Kostenfrei nachnutzbar

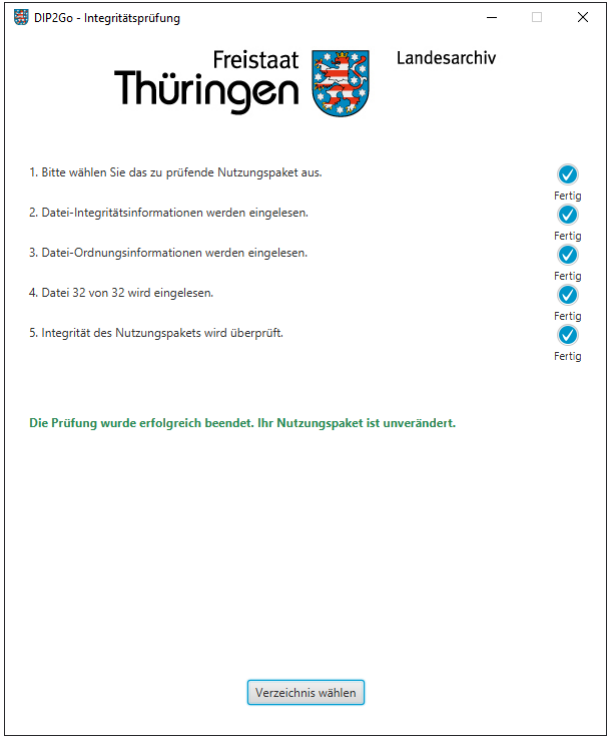
- Lizenz: LGPL
  - OpenSource-Lizenz → freie Nutzung für jeden erlaubt
  - Erfordert Zugänglichmachung von geändertem Quellcodes an den Endnutzer
  - erlaubt Einbindung auch in proprietäre Software ohne dass diese ebenfalls offengelegt werden muss
- Verfügbarkeit:
  - über den GitHub-Account des Landesarchivs Thüringen

Bereitgestellte Komponenten

- Java-Bibliothek
  - kapselt alle Funktionen die zum Erstellen und Prüfen der Integritätsdateien benötigt werden
  - Erstellt die Integritätsinformationen:
    1. Textdatei in der die Dateireihenfolge festgehalten wird
    2. Textdatei mit allen weiteren Informationen für die Integritätsprüfung (Hashwert von der Wurzel der Hausbäume, Anzahl Hashbäume, Zeitstempel …)
- Kommandozeilen-Anwendung
  - Ermöglicht das Erstellen und Prüfen der Integritätsinformationen von Nutzungspaketen
  - Kann nur Wurzeln oder kompletten Hash-Baum Speichern und Prüfen
- GUI
  - Prüft Nutzungspakete mit enthaltenen Integritätsinformationen auf Integrität und Vollständigkeit
  - Soll Nutzern für selbstständige Prüfung bereitgestellt werden

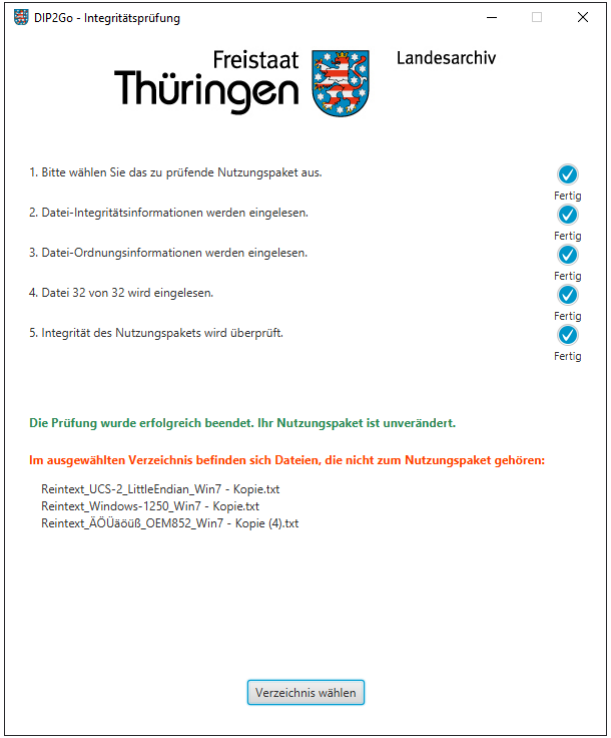
Live Demo

- Die folgenden Folien stellen die Live Demo der entwickelten Komponenten dar
- Einige der dargestellten Funktionen konnten erst im Nachhinein zum Vortrag fertiggestellt werden. Diese wurden entsprechend markiert.



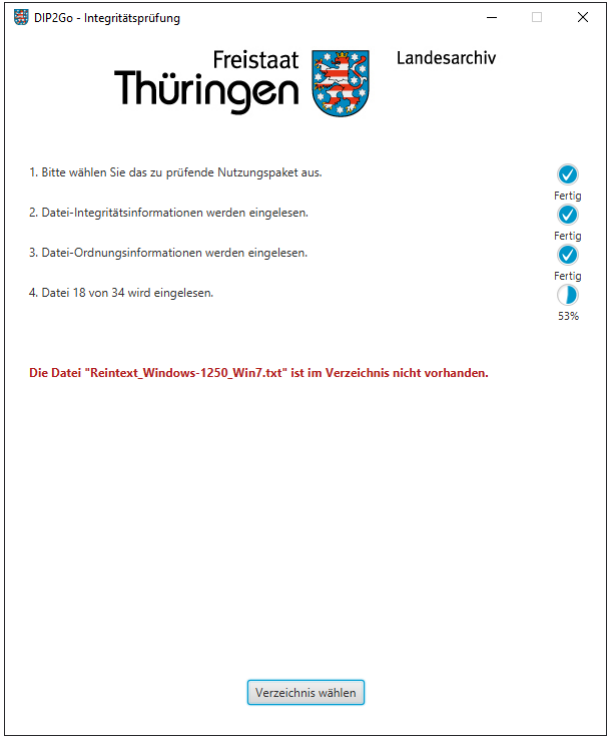
- Integritätsprüfung besteht aus folgenden Schritten:
  - Nutzungspaket auswählen
  - Datei-Integritätsinformationen einlesen
  - Datei-Ordnungsinformationen (Reihenfolge) einlesen
  - Primärdatein einlesen
  - Integrität des Nutzungspaket mittels Vergleich von Hash-Bäumen überprüfen
- Nur wenn alle Teilschritte erfolgreich waren, kann die Integrität des Nutzungspakets bestätigt werden
- Es können beliebig viele Nutzungspakete nacheinander getestet werden

- Sollte die Integritätsprüfung fehlschlagen wird eine entsprechende Fehlermeldung ausgegeben.
- Wenn bei der Erstellung der Integritätsinformationen die kompletten Hashbäume gespeichert wurden, können auch die Primärdateien aufgelistet werden, die nicht ihrer Originalversion entsprechen.
- Die Auflistung der veränderten Dateien wurde erst im Nachgang zum eigentlichen Vortrag fertiggestellt.



- Wenn im Nutzungspaket Dateien vorhanden sind, die nicht im ursprünglichen Nutzungspaket vorhanden waren, wird eine entsprechende Warnung angezeigt und die Dateien aufgelistet.

- Die Warnung wurde erst im Nachgang zum eigentlichen Vortrag fertiggestellt.



- Weitere Fehler die während der Integritätsprüfung auftreten können:
  - Primärdatei nicht vorhanden (siehe Abbildung)
  - Integritäts- bzw. Ordnungsdatei nicht vorhanden
  - Prüfsumme der Integritäts- bzw. Ordnungsdatei ist nicht korrekt
  - Schema der Integritäts- bzw. Ordnungsdatei ist nicht korrekt
- Für alle der genannten Fehler werden entsprechende Fehlermeldungen angezeigt.

Kontakt

Tony Grochow

Landesarchiv Thüringen
Digitales Magazin
Marstallstr. 2
99423 Weimar

03643/870-267
tony.grochow@la.thueringen.de

Konrad Meckel

Landesarchiv Thüringen
Digitales Magazin
Marstallstr. 2
99423 Weimar

03643/870-178
konrad.meckel@la.thueringen.de