

Automatisierte Qualitätssicherung von Retrodigitalisierungs- projekten - Revisited

8. Workshop Retrodigitalisierung, Berlin
Tony Franzky, 20.03.2026

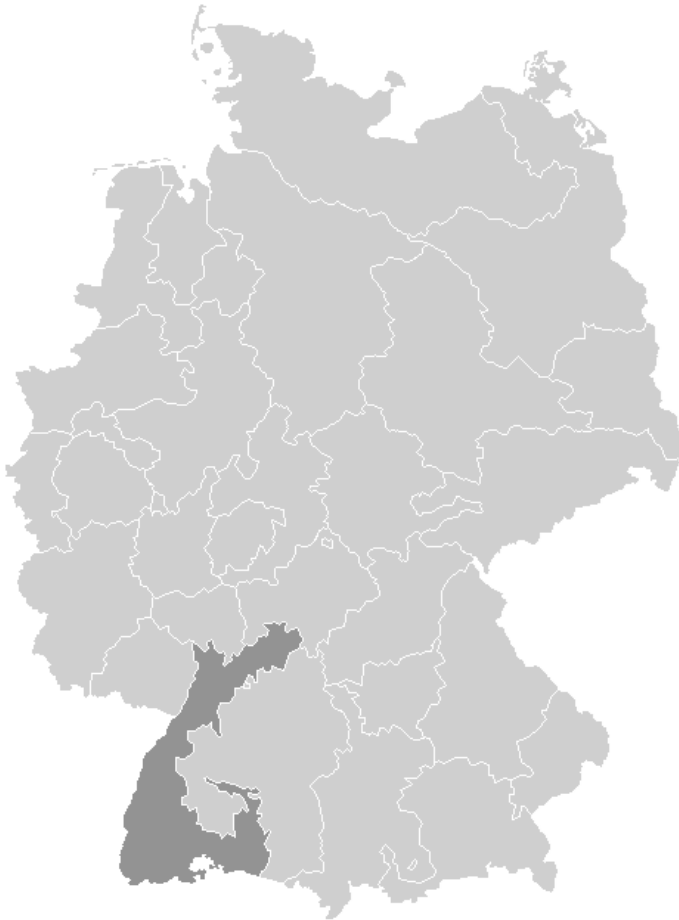
CC BY-SA 4.0 

Diese Datei wurde unter der Lizenz
„Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen“
in Version 4.0 (abgekürzt „CC-by-sa 4.0“) veröffentlicht.

<https://creativecommons.org/licenses/by-sa/4.0/>

Inhalt

- Das Erzbischöfliche Archiv
 - Allgemein
 - Digitalisierungsprojekte
- Digitalisierungskonzepte in a nutshell
- Was heißt Qualitätssicherung?
- Blurry: Was bisher geschah
- Blurry: Was noch geschehen wird



- 1825 / 27 gegründet
- Bischofssitz in Freiburg
- **1049** Pfarrgemeinden in 224 Seelsorgeeinheiten
- **29.000** Mitarbeitende
- **1,8 Mio** Katholik:innen

- Archivneubau 2003
- **11km** Akten, zwei Außenstellen in Sigmaringen und Eberbach
- **15** Personen
- Bestände aus knapp **1000 Jahren**



Digitalisierungsprojekte im EAF

Archivalien

- **Conrad Gröber** (216 Faszikel, 102.425 Scans + Derivate, 3,95 TB)
- **Maria Gröber** (10 Faszikel, 1581 Scans + Derivate, 48,7 GB)
- **B2/NS** (183 Faszikel, 88.036 Scans + Derivate, 3,23 TB)
- **Urkundenbestände** (ca. 5500 Urkunden, 8709 Scans + Derivate, 1,53 TB), *andauernd*
- **Kirchliche Amtsblätter (1)** (LZ 1970-2010, 10.295 Scans)
- **Kirchliche Amtsblätter (2)** (LZ 1853-1969, 17.245 Scans + Derivate, 42,8 GB)
- **Fotoglasplatten** (2427 Scans, 30GB), *andauernd*
- **Kirchenbücher** (14.763 Bücher, 3.1 - 3.7 Millionen Scans, 20 TB - 50 TB), *andauernd*

Extern

- **Bauaktenbestand** (4229 Scans, 550 GB), extern, *andauernd*
- **Planarchivdaten** (ca. 73.503 Scans + Derivate, 7,49 TB), extern, *andauernd*

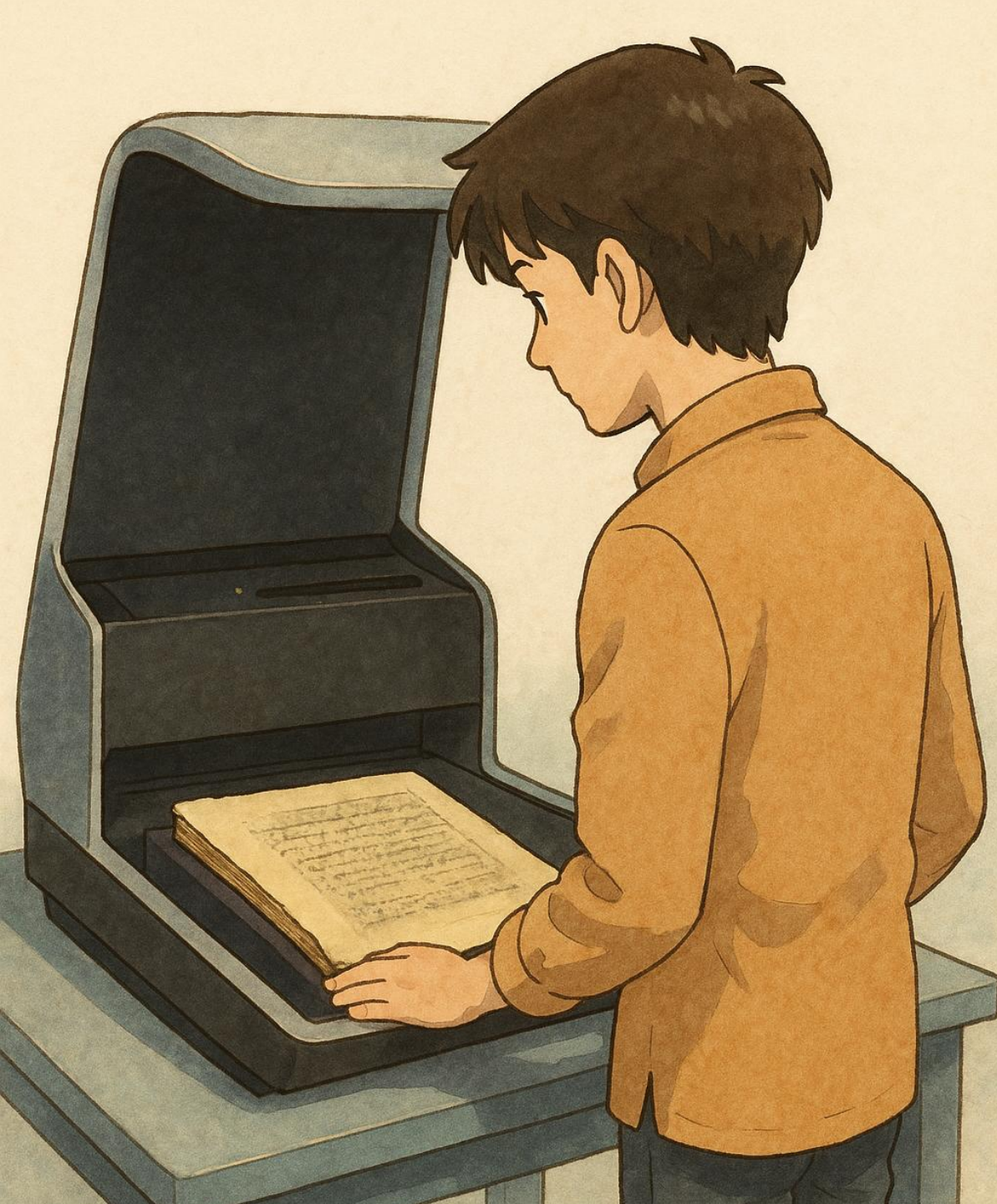
On Hold

- **Gemischter Sammelfotobestand** (ca.18.267 Durchlichtmedien), *geplant*
- **Predigtensammlung** (ca. 1000 Kassetten, 200 GB - 1000 GB), *geplant*
- **Fotonachlass Fehrenbach** (ca. 3440 Durchlichtmedien und Kontaktabzüge), *geplant*
- **Nachlass Ginter**, *geplant*

Findmittel

- **Priesterkarteikarten** (6662 Karten, 7747 Scans, 14,3 GB)
- **Kartei Priester im Kriegsdienst** (248 Karten, 496 Scans)
- **Kartei Mikroverfilmungen Kirchenbücher** (4698 Scans, 1,23 GB)
- **Urkundenrepertorien** (2029 Scans, 600 MB)
- **Retrokonversion Repertorien** (z.B. 297 Pfarrarchivrepertorien), *andauernd*

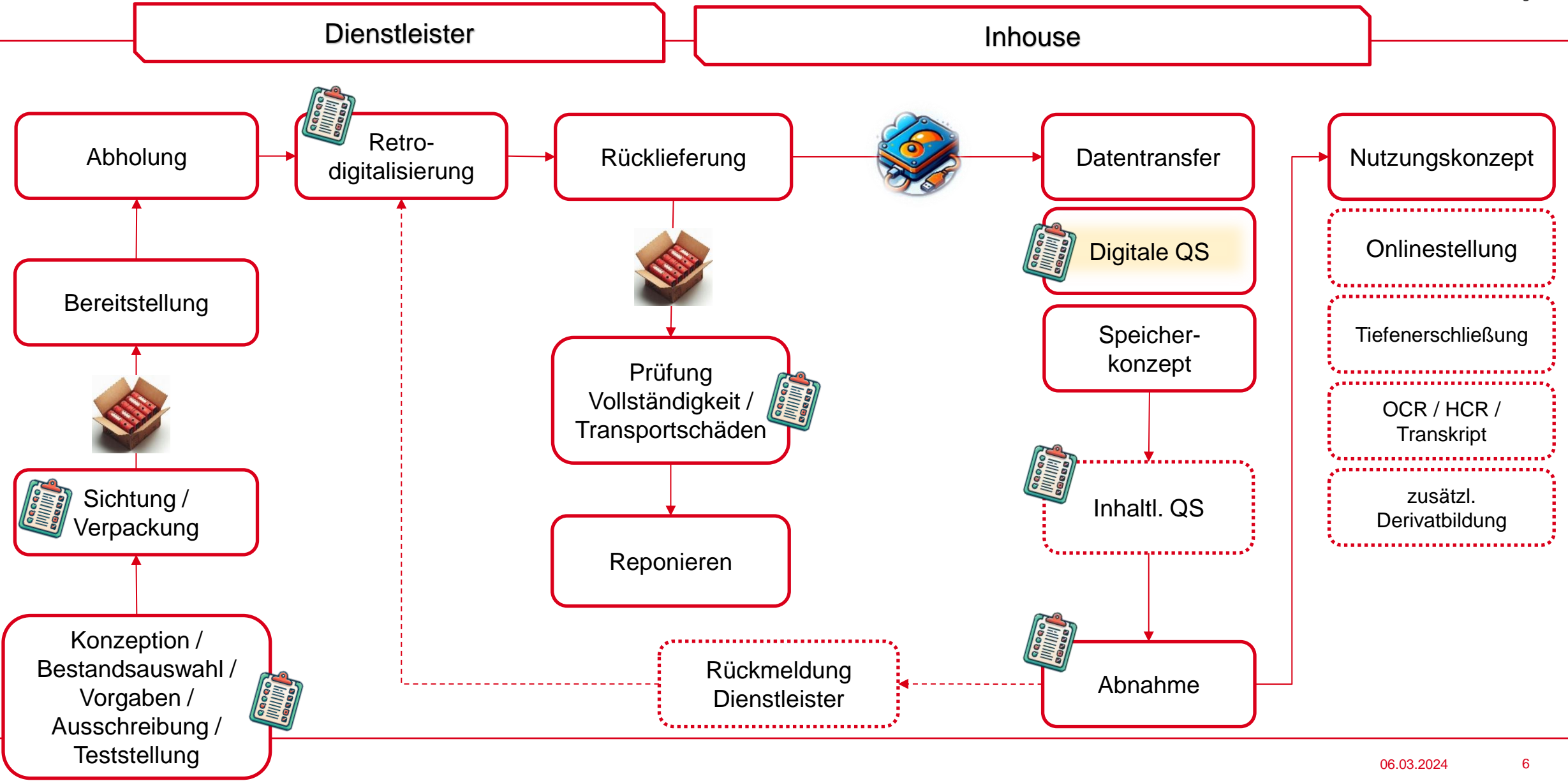
14.763 Kirchenbücher
3M+ Scans
20 TB +



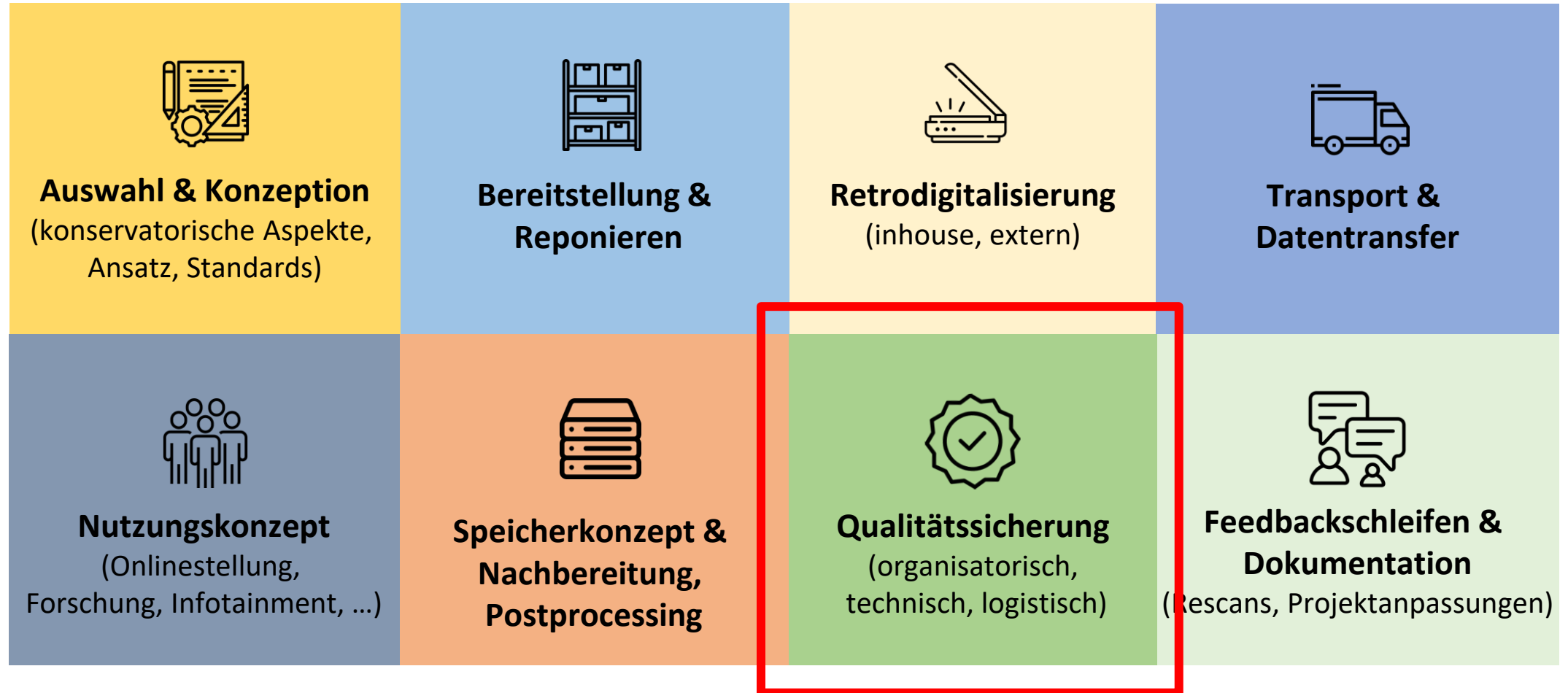
10

Digitalisate
während Vortrag





Digitalisierungskonzepte in a nutshell



Qualitätssicherungsmaßnahmen



Organisatorisch

- Scannerwahl
- Zeitplanung
- Prozesssicherung
- Inhouse / Extern Workflows



Logistisch

- Beschädigungsfall (Brand, Wasser, Physisch, ...)
- Verpackung und Transportvorgaben
- Lagerung und Klimatisierung
- Diebstahlschutz
- Versicherung



Technisch

- Scanner Kalibrierung
- IT Sicherheit
- Geräteunterweisung
- Datenauslieferung, -transfer und -güte

Verwendete Software



FileListCreator

stabil auch bei 100k+ Dateien
portable nutzbar

- Dateiinherente Metadaten auslesen (IPTC, EXIF, ID3,...)
- Auflösung bei Bildern und Ausrichtung
- Formatvalidierung über MIME Typ
- Dokumenteigenschaften lesen (Seitenzahl, Laufzeit, ...)
- Dubletten über Dateihashes
- Dateieigenschaften lesen

<https://www.sttmedia.com/filelistcreator>



AllDup

Prüfen von Dubletten mit verschiedenen Dateihashes und Bildähnlichkeitsalgorithmen (pHash, dHash, bHash,..)

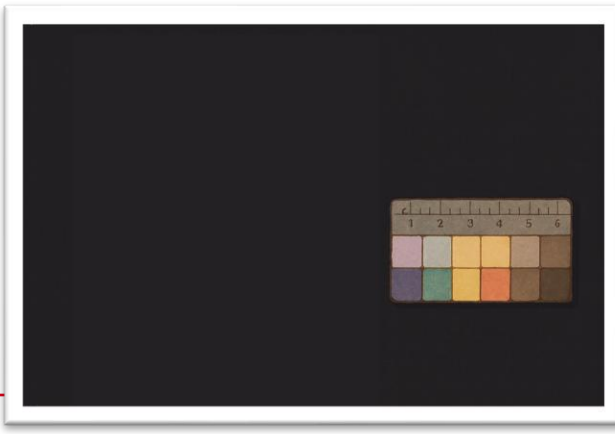
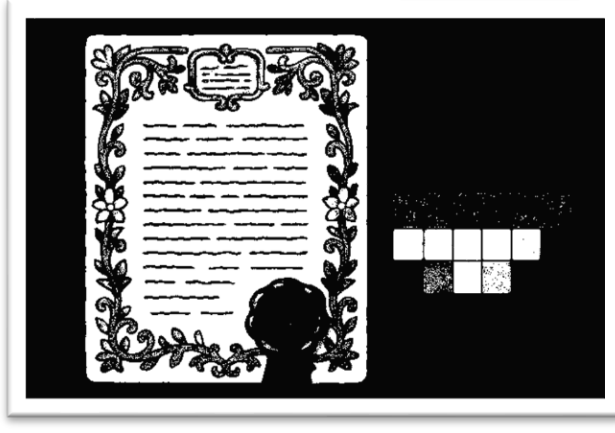
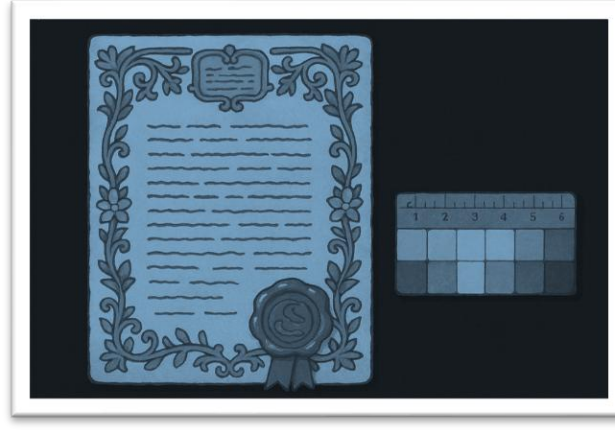
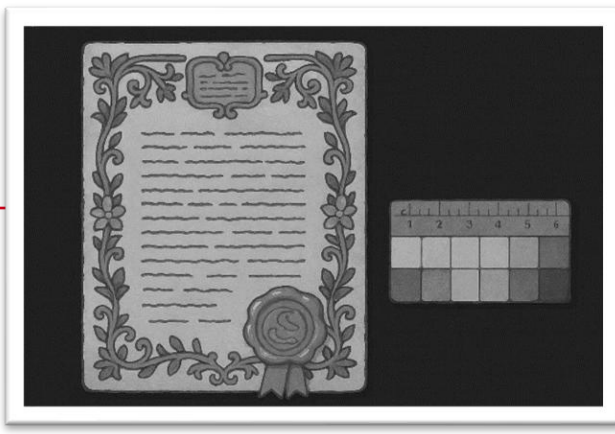
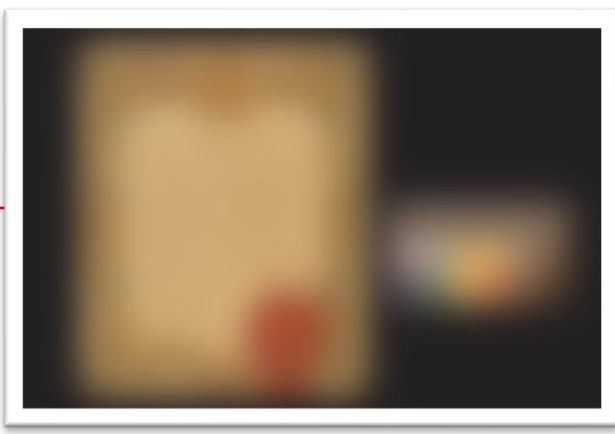
<https://www.alldup.de/>



DROID

Validierung von Dateitypen anhand der Pronom ID

<https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>



Kann man das in einem Tool automatisieren?

Erwartungshaltungsmanagement (1): Digitalisierungsbezogene Qualitätskontrolle

Strukturell

- Dateibenennung / Benennungsschema
- Datenablagestruktur
- Metadatierung

Derivate

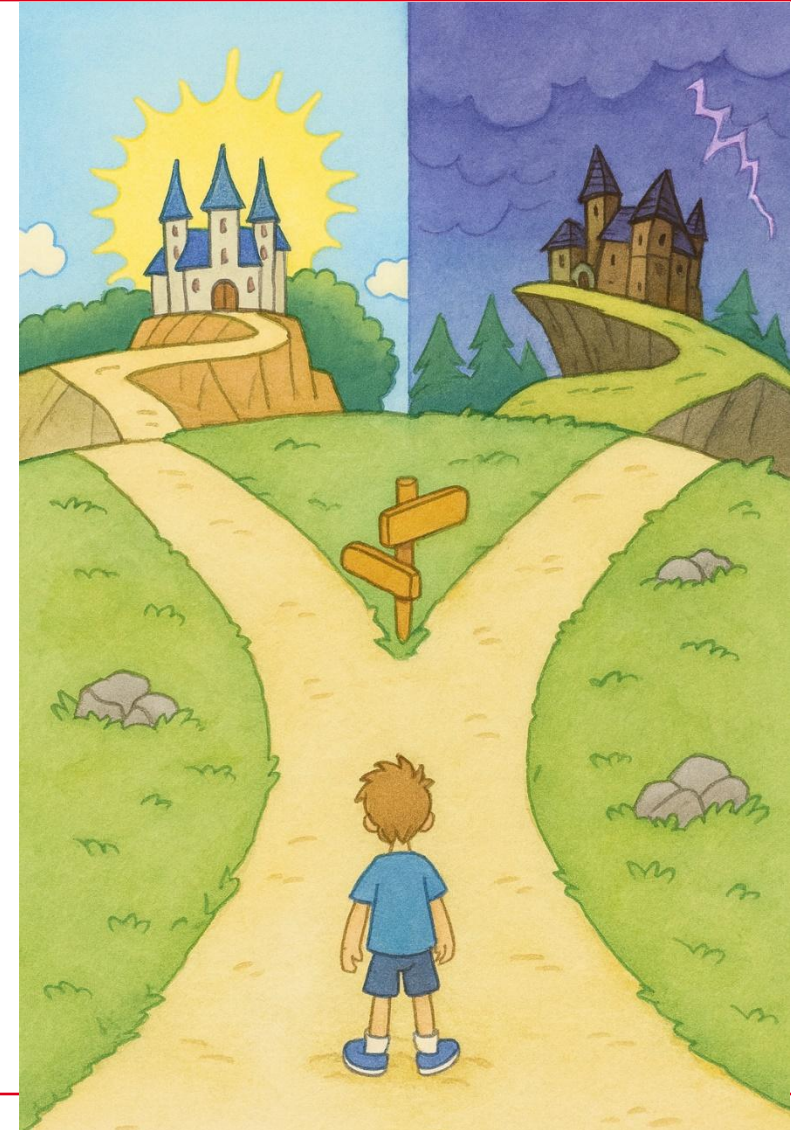
- Vollständigkeit
- Beschnitte (z.B. Farbpatches)

Dateiattribute

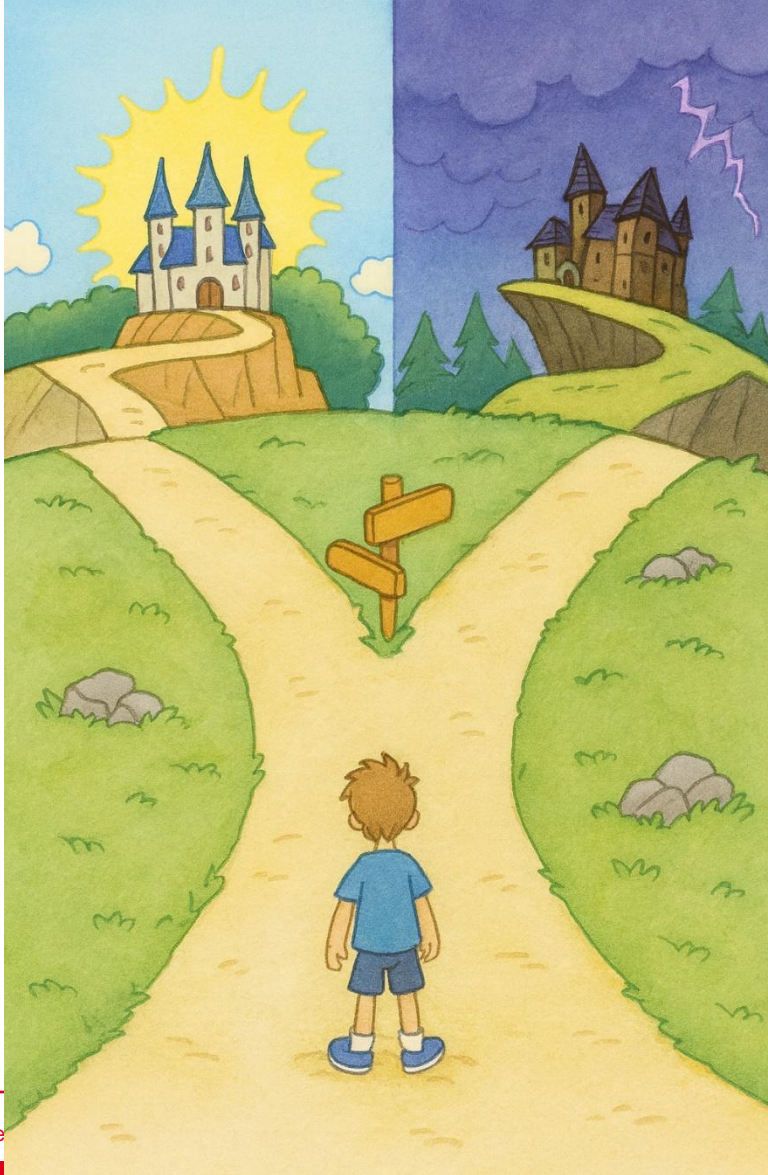
- Auflösung
- Speichergrößenvorgaben
- Farbraum (RGB, CMYK, HLS, HSV, sRGB...)
- Dateiintegrität, -konformität und -validierung
- Dateiinterpretierbarkeit

Optische Qualität

- Fremdobjekte (Finger, Kabel, ...)
- Artefakte (Staub, Kratzer, Moiré)
- Unschärfe
- Dokumentausrichtung
- Leerseiten
- gefaltete Dokumente
- Durchschlagpapier
- Doppelscans
- unvollständige Scans
- Verbliebene Trennpatches
- Beschnitt
- Tonwertwiedergabe / Farbgenauigkeit
- Bildrauschen
- Geometrische Verzeichnung
- Gleichmäßige Ausleuchtung / Abschattungen



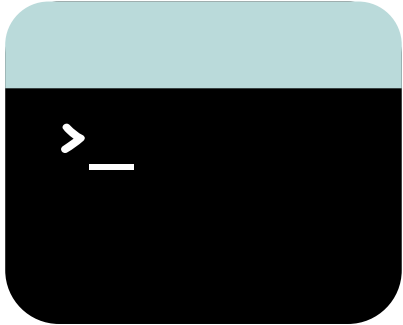
Erwartungshaltungsmanagement (2): Digitalisierungsbezogene Qualitätskontrolle



Was nicht gemeint ist:

- Text- oder Handschrifterkennung
- Bildelemente oder –inhalte Erkennung
- Automatische Verzeichnung
- Vorschaubilder, Thumbnails, Volumenkurven, Transkripte
- Irgendwas mit KI
- Informationretrieval & -extraction

Blurry: Was bisher geschah?



Stand 2024:

- Unscharfe Bilder
- S/W oder Farbigkeit
- Leerseiten

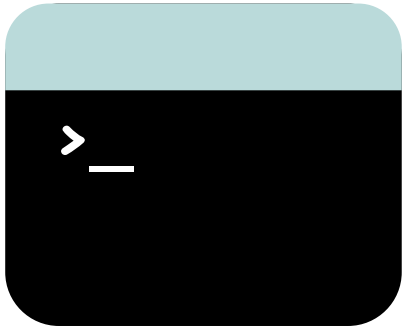
Stand 2025:

- Hashwerte berechnen (Message-Digest & Secure Hash Familie)
- PRONOM Dateisignaturen
- Exif, XMP, IPTC Metadaten
- Regex für Namensbenennung
- QR Code Reader (Trennpatches)
- Datenstrom korrupt
- Ausrichtung, Auflösung, Speichergröße
- Bildähnlichkeiten via pHash

Techstack:

Python, CMD Tool, Docker, Apache 2.0 Lizenz

Blurry: 2026



Python Modul

- zur Einbettung in eigene Pythonprojekte
- komplette Funktionsumfang für Eigenscripte, Insellösungen, Enthusiasten:innen, etc.
- Basis, andauernde Weiterentwicklung

Kommandozeile

- Einbindung in existierende Anwendungen, Workflowengines, ...
- mit Docker (BETA) und kompilierter Windowsversion (ALPHA)

StreamlineAPP (BETA)

- Handson für Endanwender:innen
- Aufsetzen in eigener Infrastruktur notwendig
- Experimentieren für Resultate
- Docker (BETA)

Technische Roadmap

- RestAPI
- VLM Streamlining (experimentell)
- „echte“ GUI
- Härtung der Dockerfiles

Tests und Kooperationen

- Implementierung
- Testung
- Lücken in der Dokumentation



Vielen Dank für Ihre Aufmerksamkeit.

Link zum Tool

<https://github.com/sugob05/blurry>

Offene Fragen?



Tony Franzky
Sachgebietsleiter digitale Langzeitarchivierung
Erzbischöfliches Archiv Freiburg

tony.franzky@ordinariat-freiburg.de

Tel.: +49 (0) 761 2188 493

Fax.: +49 761 2188 76 493