



Landesarchiv  
Baden-Württemberg

# OCR im Archiv Ein Blick in das FDMLab@LABW

**Abschlussworkshop OCR-BW, 09.06.2021**  
Benjamin Rosemann, Elisabeth Klindworth

Laufzeit von Juli 2020 bis Juni 2022

## Kontext

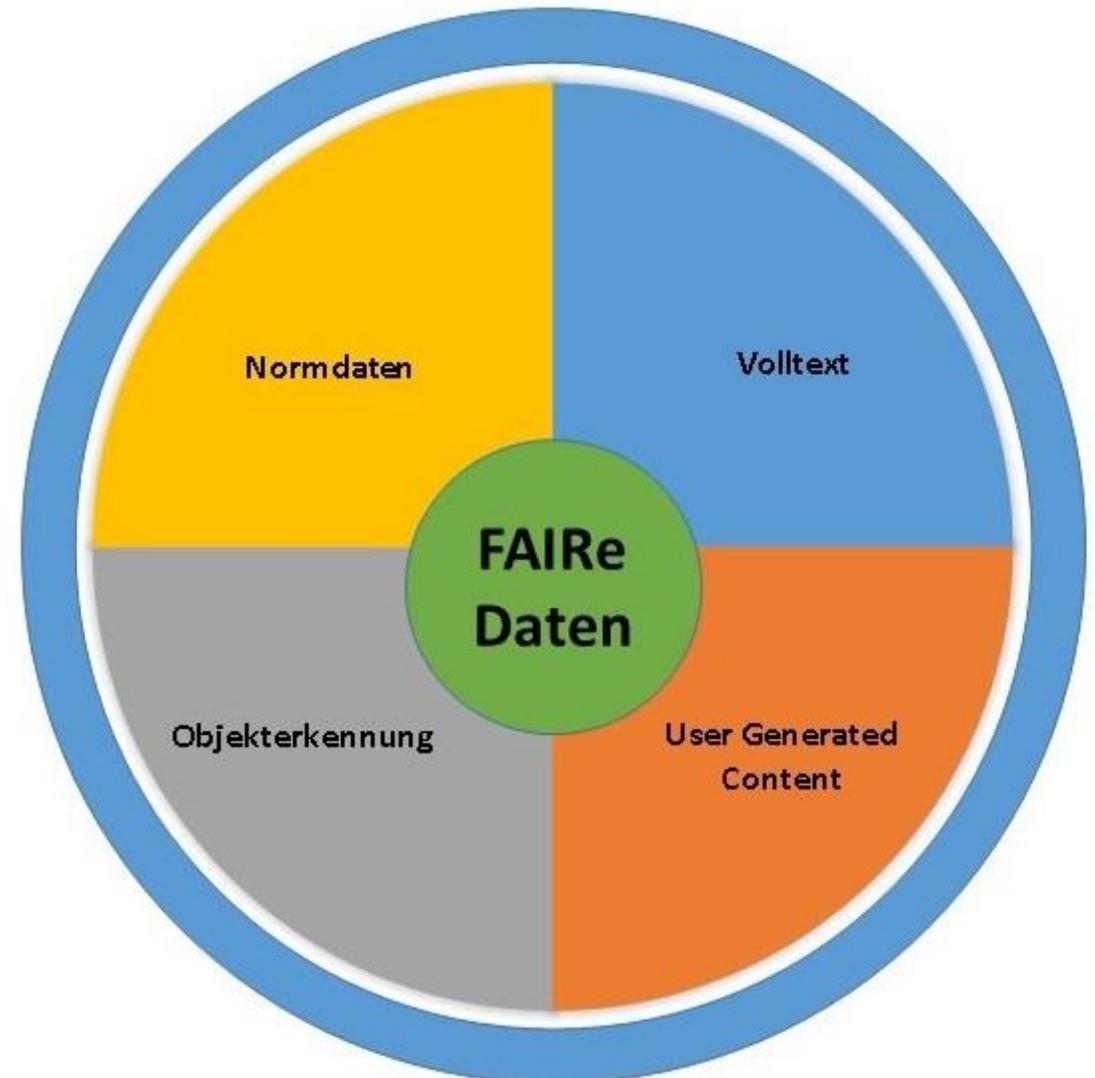
- Das Projekt steht im Kontext des Aufbaus einer nationalen Forschungsdateninfrastruktur (NFDI).
- Förderung durch die Baden-Württemberg Stiftung im Rahmen der Zukunftsoffensive III

## Projektziele

- Aufbau einer im Landesarchiv angesiedelten Basisinfrastruktur im Bereich E-Science und Forschungsdatenmanagement (FDM)
- Verbesserung der Zugänglichkeit und Nutzbarkeit des Archivguts
  - Antworten auf die steigende Nachfrage nach standardisierten, übergreifend auswertbaren Daten geben

## Weitere Informationen

Projektseite auf der LABW-Homepage [1]  
Blog mit Praxisberichten aus dem Projekt [10]





1.

# Was digitalisieren wir?

Betr. Anfrage des Gaupersonalamtes Baden, Hauptstelle für pol. Leiter  
vom 26. November 1936.

# Archivgut

Am: Bad. Rechnungshof

3. Wohnort: Karlsruhe Straße: Händelstr. 21 Konf. evang.

4. Mitglied der N.S.D.A.P. nein Seit wann? -- Mitgl. Nr. --

5. Ortsgruppe: -- Kreis: --

6. Mitglied des RDB oder einer anderen von der N.S.D.A.P. betreuten Organisation: ja  
Welcher? RDB / NSV Seit wann? 1.1.34/Mai 1934

7. Arbeitsdienst? nein Frontsoldat? ja  
Seeresdienst? nein

8. Mitglied der SA, SS usw. nein Seit wann? --

9. Freimaurer? nein von wann bis wann? --  
Loge? -- Grad? --

10. Welcher Partei hat er früher angehört, wo und wie lange? Deutsche liberale Volkspartei  
von Gründung bis 1933

11. Hat er sich früher für oder gegen die N.S.D.A.P. ausgesprochen? gegen NSDAP

12. Bejaht er den nationalsozialistischen Staat? nein

13. Ist er in der Lage im nat. Sinne erzieherisch auf seine Volksgenossen einzuwirken? nein

14. 5 Ist er Begleiter der NS-Presse? nein Karlsruher Tagblatt  
Benjamin Rosemann | Elisabeth Klindworth

15. Ist er regelmäßiger Besucher der Schulungs- und Kameradschaftsabende, sowie der Versammlungen und

## Merkmale unseres Archivguts:

- komplexe Layouts (Formulare, Tabellen)
- Mischung aus Hand-, Druck- und Schreibmaschinenschrift
- unleserliche Handschriften
- Akten mit vielen Dokumententypen
- z. T. Abnutzungserscheinungen und Schäden am Papier
- Teilweise Mikrofilme als Grundlage

1.

Stuttgart, den 26. April 1879.

**AN DEN KOENIG**

Anbringen

D. Dr. de Sarwey.

des Staats-Ministeriums.

Entwurf und Ausführung eines Königlichen Verordnungs wegen Einsetzung eines Kommissars für Gesundheitswesen bei dem künftigen Landtag in Stuttgart.

Euer Königlich-Majestät

haben sich im Auftrage des und von dem Landtag dem Landtag mit Zustimmung des Landesparlamentes, bestehend aus den Mitgliedern des Landesparlamentes, die Einsetzung eines Kommissars für Gesundheitswesen bei dem künftigen Landtag in Stuttgart und dem unter dem Vorsitz des Landtagspräsidenten verhaltenen, Euer Königlich-Majestät erlassen, geneigt geneigt geneigt hat die Ausführung Koethe's des Kommissars, Kallab's des Landtagspräsidenten, bestehend aus dem

St. d. 29. April 1879.  
54.

8/S.

Landesarchiv BW, HStA E 33 Bü 354, Bild 24

No. 106. H. Sitzung 5.5.79

Landtag - Protokoll No. 122.

Geneigt und geneigt, Protokollbestätigung und, wenn möglich, Fortsetzung, jedoch kein Recht, ungeschaffen zu sein, ist in dem Protokoll und der Fortsetzung zu thun.

ggg. v. Krichig

Ein König haben wir Kraftprotokoll, Abgabe zu machen und das dem Landtag zu thun.

ggg. Haller.

**Namentliches Verzeichnis.**

Landtagsnummer	Titel	Name	Seite	Bemerkungen
1	Landtag	Friedrich Karl, Herr	12	
2	Landtag	Herr, Herr	13	
3	Landtag	Herr, Herr	14	
4	Landtag	Herr, Herr	14	
5	Landtag	Herr, Herr	16	
6	Landtag	Herr, Herr	16	
7	Landtag	Herr, Herr	18	
8	Landtag	Herr, Herr	18	
9	Landtag	Herr, Herr	20	
10	Landtag	Herr, Herr	20	
11	Landtag	Herr, Herr	22	Am 15. 18. 1879 ist Landtag in Stuttgart.
12	Landtag	Herr, Herr	22	
13	Landtag	Herr, Herr	24	Am 16. 18. 1879 ist Landtag in Stuttgart.
14	Landtag	Herr, Herr	24	
15	Landtag	Herr, Herr	26	
16	Landtag	Herr, Herr	26	
17	Landtag	Herr, Herr	28	
18	Landtag	Herr, Herr	28	
19	Landtag	Herr, Herr	30	
20	Landtag	Herr, Herr	30	
21	Landtag	Herr, Herr	32	
22	Landtag	Herr, Herr	32	
23	Landtag	Herr, Herr	34	
24	Landtag	Herr, Herr	34	
25	Landtag	Herr, Herr	36	
26	Landtag	Herr, Herr	36	
27	Landtag	Herr, Herr	38	
28	Landtag	Herr, Herr	38	
29	Landtag	Herr, Herr	40	
30	Landtag	Herr, Herr	40	

Landesarchiv BW, HStA M 438 Nr. 122, Bild 6

- 1 8 (A 187) 1605–1608 (1605)  
2 Bürgermeister und Rat der Stadt **Aalen**  
. / .  
3 Johann Christoph, Propst zu Ellwangen, Kanzler und Räte das., Wiguläus von Erolzheim, ellwangischer Amtmann zu Kochenburg  
4 Kl.: Dr. Georg Amandus Wolff 1605 (1604)  
Bekl.: Dr. Johann Jacob Kölblin (1605)  
5 secundi mandati der Pfändung cum mandato de non offendendo

# Findbücher

- 7 Q 3 Instr. des ksl. Notars und württ. Oberratssekretärs M. Christophorus Schmidlin über in Stuttgart vollzogene Protestation des Aalener Rats 1605  
Q 6/7 Verträge zwischen Ellwangen und Aalen betr. Zehnten auf Aalener Flur 1587  
Q 8–13 Urfehden und Gelöbnisse Aalener Bürger 1605  
ohne Q Zwischenurteil und weitere Bescheide des Lehengerichts Ellwangen 1607  
8 Q 1–13 (ohne 4), 8 a–11 (= teilweise Doppelzählung) und 6 weitere Schriftstücke  
3 cm

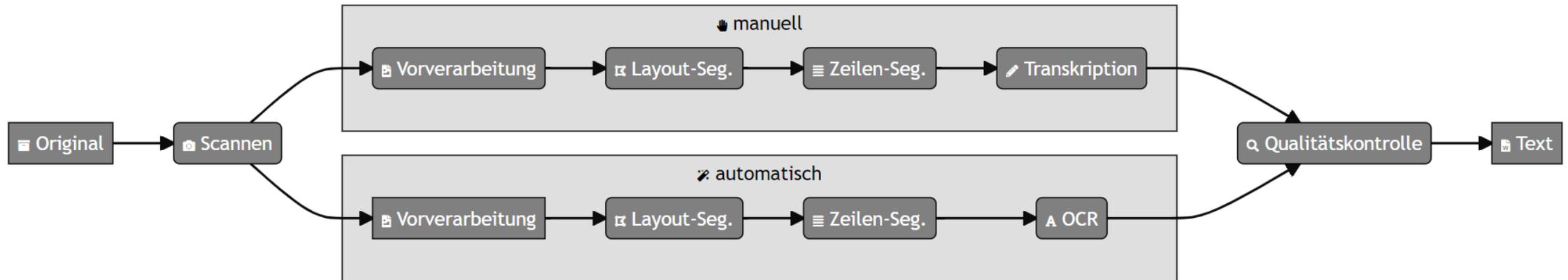
- 1 9 (A 188) 1617–1625  
2 Bürgermeister und Rat der Stadt **Aalen**  
. / .  
3 Johann Christoph, Propst zu Ellwangen  
4 Kl.: Lic. Johann Peter Mörder 1617 (1614)  
Dr. Christoph Stauber 1620  
Bekl.: Dr. Johann Jakob Kölblin 1617 (1613)  
5 (tertii) mandati auf die Konstitution der Pfändung  
Verstrickung des Aalener Gerichtsprokurators Caspar Lentlin wegen von klag. Stadt angeordneter Haussuchung im Weiler Himmlingen. Stift Ellwangen beansprucht dort, wobei Vogtei und niedere Gerichtsbarkeit Aalen zugestanden wird, hohe Obrigkeit.  
7 Q 5, 13 Urfehden des Jacob Weng und Philipp Tollinger 1484, 1585  
Q 10 eigenhändiger Bericht des Wilhelm Krauß, Stadtschultheiß zu Aalen, betr. Ehe- und Inzestsache des Jacob Weng, Himmlingen 1618  
Q 17/18 Urfehde des Lienhard Harsch, Sohn des Konrad Harsch, Ziegler von Himmlingen, 1465 (Ausf. U 1 und Abschr.) sowie des Matthias Spiegel von Vaihingen, Conz Schürger von Crailsheim und Utz Lay von Bergzabern (Ausf. und Abschr.) 1560  
8 Q 1–18  
3 cm  
Protokoll beschädigt

Neben Archivgut werden auch Findbücher digitalisiert, um die Erschließungsdaten über das Online-Findmittelsystem (OLF) des LABW recherchierbar zu machen.



**2.**

**Welche Tools setzen wir ein?**



## manuell

- OCR4All [3]
- Transkribus [4]

## automatisch

- Abbyy FineReader
- OCR-D [5]
- Transkribus [4]

## Qualitätskontrolle

- dinglehopper (Qurator) [6]
- PAGE Viewer (PRImAResearch) [7]

# Quality metrics for OCR texts

## Metrics:

- Character Accuracy (CA = 1 - CER)
- Word Accuracy (WA = 1 - WER)
- Bag of Characters (BoC)
- Bag of Words (BoW)
- Flexible Character Accuracy (FCA) [8]

**Errors:** Accuracy = 1 - ErrorRate  
= 1 - Errors / Chars<sub>GT</sub>

	GT	OCR	Weight		
Delete	ab	a	1	1	1
Insert	a	ae	1	1	0
Replacement	a	e	1	2	1

## Example:

1	2	3
4	5	6
7	8	9

row-wise: 1 2 3 4 5 6 7 8 9

col-wise: 1 4 7 2 5 8 3 6 9

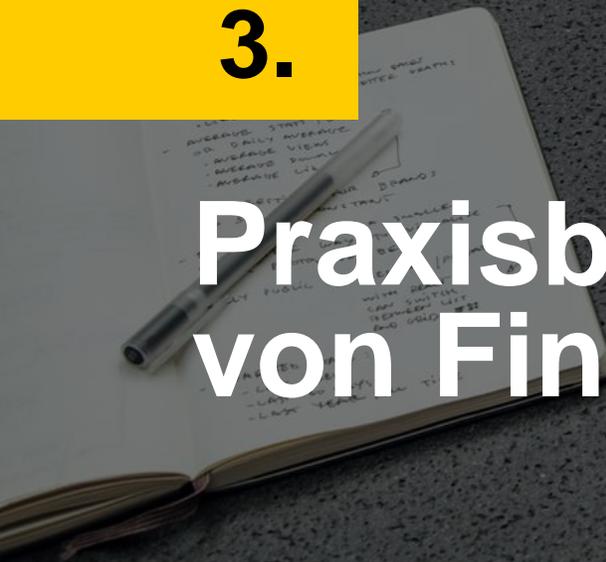
## Properties:

	CA	WA	BoC	BoW	FCA
Order					
Speed					

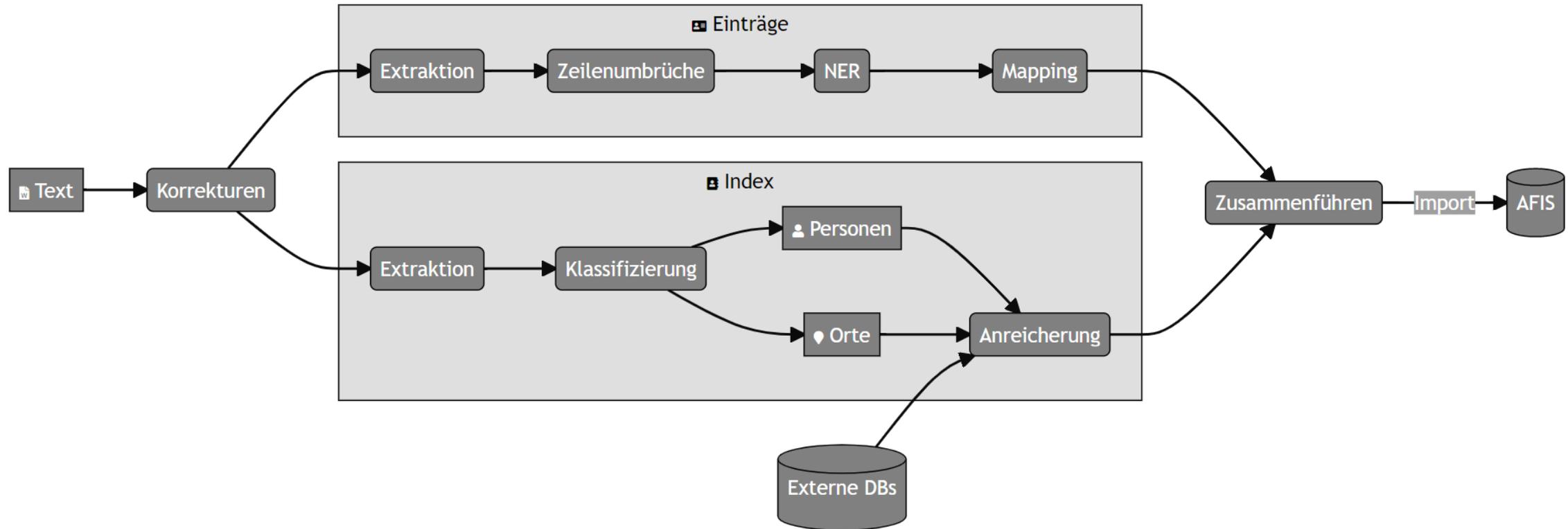


3.

# Praxisbeispiel: Digitalisierung von Findbüchern



# Workflow Digitalisierung Findbücher



**Abgeschlossen:** Findbuch zu Urkunden der Grafschaft Friedberg-Scheer (ca. 1 500 Urkunden) [9]

**In Arbeit:** Findbücher zu Akten des Reichskammergerichts (ca. 10 000 verzeichnete Einheiten)

- Manuelle Nacharbeiten: Unregelmäßigkeiten bei Struktur, Layout und Sprache
- Analoges Findbuch stets zur Hand haben: Fehler in digitaler Version ausbessern
- Named Entity Recognition nicht ausreichend
- Extraktion transparent gestalten z.B. via Zwischenformat zur Visualisierung
- Enge Zusammenarbeit von technischem und archivfachlichem Personal
  - Granularität der Datenextraktion gemeinsam bestimmen
  - Mapping der extrahierten Daten auf Erfassungsformular abstimmen



**Vielen Dank!**

**Benjamin Rosemann**  
Landesarchiv Baden-Württemberg  
Zentrale Dienste  
Projekt FDMLab@LABW  
URL: <https://fdmlab.landesarchiv-bw.de>  
Tel.: +49 711 335075-512  
E-Mail: [benjamin.rosemann@la-bw.de](mailto:benjamin.rosemann@la-bw.de)

**Elisabeth Klindworth**  
Landesarchiv Baden-Württemberg  
Archivischer Grundsatz  
Projekt FDMLab@LABW  
URL: <https://fdmlab.landesarchiv-bw.de>  
Tel.: +49 711 335075-513  
E-Mail: [elisabeth.klindworth@la-bw.de](mailto:elisabeth.klindworth@la-bw.de)

1. Projektseite FDMLab@LABW: <https://www.landesarchiv-bw.de/de/landesarchiv/projekte/fdmlab%2540labw-/71653>
2. NS Überlieferung in staatlichen Archiven: <https://www.leo-bw.de/themenmodul/sudwestdeutsche-archivalienkunde/querschnittsartikel/ns-uberlieferung>
3. OCR4All: <http://www.ocr4all.org/>
4. Transkribus: <https://transkribus.eu/>
5. OCR-D: <https://ocr-d.de/>
6. Dinglehopper: <https://github.com/qurator-spk/dinglehopper/>
7. Page Viewer: <https://www.primaresearch.org/tools/PAGEViewer>
8. Flexible character accuracy: <https://doi.org/10.1016/j.patrec.2020.02.003>
9. Findbuch Staatsarchiv Sigmaringen Dep. 30/1 T 1, Grafschaft Friedberg-Scheer: <https://www2.landesarchiv-bw.de/of21/olf/struktur.php?bestand=2240>
10. Blog mit Praxisberichten aus dem FDMLab@LABW: <https://fdmlab.landesarchiv-bw.de>