

Digitalisierungsaktivitäten und Kitodo an der UB Mannheim

Kitodo Praxistreffen 2024 in Marburg

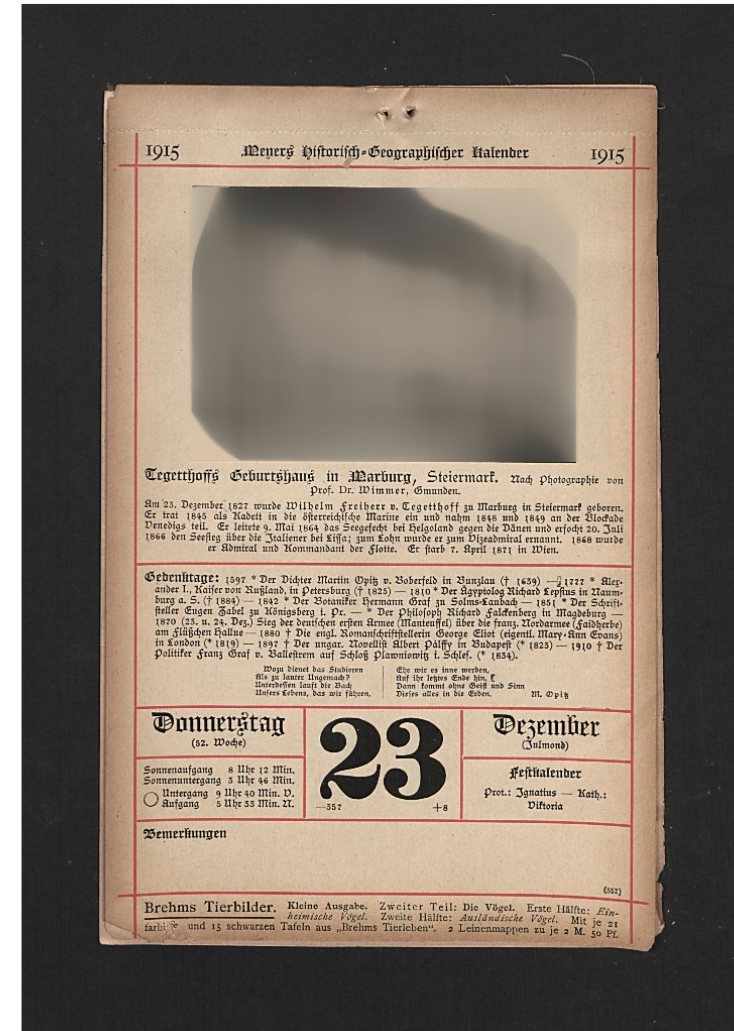
Datenlieferung für Deutsche Digitale Bibliothek

- Das von der DFG geförderte Digitalisierungsprojekt *Deutscher Reichsanzeiger und Preußischer Staatsanzeiger* (1871–1945) liefert rund 20000 METS/MODS-Dateien (XML) an das Deutsche Zeitungsportal.
- Idealerweise sind alle Daten fehlerfrei und können vollständig übernommen werden.
- Der Shellskript *validate-mets* prüft, ob die XML-Dateien fehlerfrei (valide) sind und ob alle Vorgaben der DDB (Anwendungsprofil) eingehalten werden.
- Nicht nur METS/MODS-Dateien für Zeitungen, sondern auch für andere digitalisierte Medien können von *validate-mets* geprüft werden. Dafür gibt es eigene Vorgaben.

<https://github.com/Deutsche-Digitale-Bibliothek/ddb-metadata-schematron-validation>
<https://github.com/UB-Mannheim/Reichsanzeiger/tree/main/scripts/validate>

Bereitstellung (noch) rechtlich bewehrter Objekte

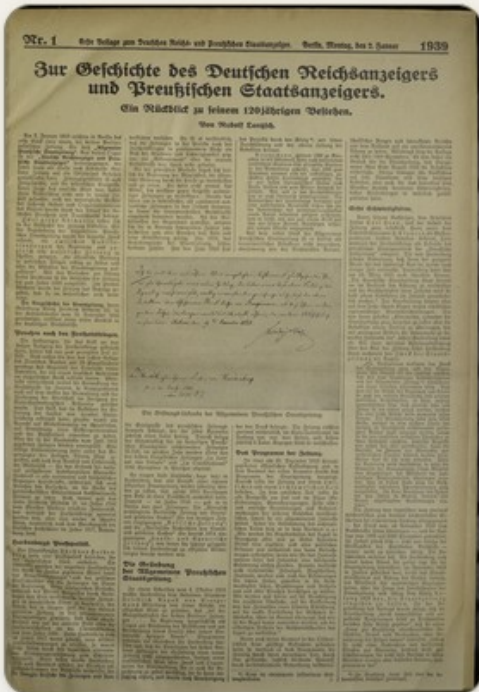
- Gemeinsamer DFG-Projektantrag von SLUB Dresden, SUB Hamburg und UB Mannheim für „Entwicklung von Werkzeugen zur rechtssicheren Präsentation multimedialer Objekte mit Kitodo.Presentation“
- Beispiel aus Mannheim: urheberrechtlich geschützte Fotos in Meyers Historisch-Geographischer Kalender
- Lösung: Koordinaten der Fotos werden automatisiert bestimmt für Berechnung neuer Ersatzbilder



Volltexte für Kitodo mit eScriptorium

- Transkriptionsplattform eScriptorium läuft an der UB Mannheim seit drei Jahren.
- Inzwischen mehr als 350 registrierte Nutzer*innen, drei Instanzen (öffentliche Produktivinstanz mit Softwarestand 2023, Clone der Produktivinstanz mit aktuellem Softwarestand, interne Testinstanz).
- Texterkennung für Kitodo.Presentation (bei Handschriften oder schlechter OCR)
 - Import der Bilddateien in eScriptorium per URL auf METS XML
 - Segmentierung und Texterkennung in eScriptorium, in der Regel mit manuellen Korrekturen und Nachtraining
 - Export von ALTO XML aus eScriptorium für Volltextanzeige in Kitodo.Presentation

Claude.ai – LLMs als Zukunft der Texterkennung?



Der Text ist ein Rückblick auf die Geschichte des Deutschen Reichsanzeigers und des Preussischen Staatsanzeigers anlässlich ihres 120-jährigen Bestehens im Jahr 1939. Er beschreibt die Gründung dieser offiziellen Veröffentlichungsorgane, ihre Entwicklung und Bedeutung für die Verbreitung staatlicher Informationen und Gesetze im Deutschen Reich und in Preußen. Außerdem geht es um die Herausforderungen bei der Produktion und Auslieferung der Zeitungen in früheren Zeiten. Der Autor betont die Wichtigkeit einer zuverlässigen staatlichen Veröffentlichung für die Rechtssicherheit und Transparenz. Der Text vermittelt einen historischen Überblick über diese zentralen Publikationen im Kontext der damaligen staatlichen Strukturen.

S Bitte fasse den Text zusammen.

Aus: Weil, S., & Gottschling, T. (2024).
Fünf Jahre DFG-Projekt Deutscher Reichsanzeiger – was haben wir gelernt?
112. BiblioCon 2024, Hamburg, Germany.
<https://nbn-resolving.org/urn:nbn:de:0290-opus4-192038>

OCR für einfache Texte mit lokalem LLM fast gut

https://digi.bib.uni-mannheim.de/fileadmin/gv/497570904_19460001/default/497570904_0007.jpg

Inhaltsverzeichnis

	Seite
Vorwort	5
Allgemeine Übersichten	6–14
I. Fläche, Gemeinden, Haushaltungen und Einwohner der Länder und Kreise der sowjetischen Besatzungszone	6
II. Die Gemeinden mit 10000 und mehr Einwohnern in der sowjetischen Besatzungszone (nach der Größe geordnet)	14
Systematisches Verzeichnis der Gemeinden	15–66
Land Brandenburg	16
Land Mecklenburg	25
Land Sachsen-Anhalt	35
Land Thüringen	46
Land Sachsen	56
Alphabetische Verzeichnisse	67–104
I. der Gemeinden	68
II. der Kreise	104
Karte: Die Stadt- und Landkreise der sowjetischen Besatzungszone (Stand 1. Januar 1948)	

****Inhaltsverzeichnis****

- * ****Vorwort****
- * Seitenzahl: 5
- * ****Allgemeine Überlegungen****
- * Seitenzahlen: 6-14
- * ****Systematische Verzeichnisse der Gemeinden****
- * Seitenzahlen: 15-66
- * Land Brandenburg
- * Seite 16
- * Land Mecklenburg
- * Seite 25
- * Land Sachsen-Anhalt
- * Seite 35
- * Land Thüringen
- * Seite 46
- * Land Sachsen
- * Seite 56
- * ****Alphabetische Verzeichnisse****
- * Seitenzahlen: 67-104
- * I. der Gemeinden
- * Seite 68
- * II. der Kreise
- * Seite 104
- * ****Karte: Die Städte und Landkreise der sowjetischen Besatzungszone (Stand 1. Januar 1948)****
- * Seitenzahl: 105

****Fußzeile****

- * Stand: 1. Januar 1948

Komplexe Vorlagen von KI weniger gut erkannt

Allgemeine

I. Fläche, Gemeinden, Haushaltungen und Einwohner der nach der Volkszählung

(Gebietsstand vom

Lfd. Nr.	Gebiet	Fläche in qkm	Zahl der		Wohnbevölkerung			Auf 1 qkm kommen Einw.	wenige 200	
			Ge- mein- den ¹⁾	Haus- hal- tungen	ins- gesamt	davon			Ge- meinden	b
						männ- lich	weib- lich			
		1	2	3	4	5	6	7	8	
Länder:										
I	Brandenburg	26 976,42	2 208	867 556	2 527 492	1 066 405	1 461 087	93,7	2 019	1
II	Mecklenburg	22 937,78	2 418	633 167	2 139 640	912 385	1 227 255	93,3	2 303	1
III	Sachsen-Anhalt	24 668,80	2 660	1 378 955	4 160 539	1 801 800	2 358 739	168,7	2 321	1
IV	Thüringen	15 598,33	2 382	939 319	2 927 497	1 262 326	1 665 171	187,7	2 173	1
V	Sachsen	16 991,92	2 581	2 005 121	5 558 566	2 336 630	3 221 936	327,1 ¹⁾	2 169	1
Se.	Sowj. Besatzungs- zone insgesamt .. (ohne Berlin)	107 173,25	12 249	5 824 118	17 313 734	7 379 546	9 934 188	161,5	10 985	5

OI llama3.2-vision:latest

Allgemeine

I. Fläche, Gemein

Gebietstand vom

1. Land Brandenburg

Länd

Brandenburg (Mecklenburg)

Sachsen-Anhalt

OI llama3.2-vision:latest
Allgemeine

I. Fläche, Gemeinden, Haushaltungen und Einwohner der nach der Volkszählung
Gebietstand vom

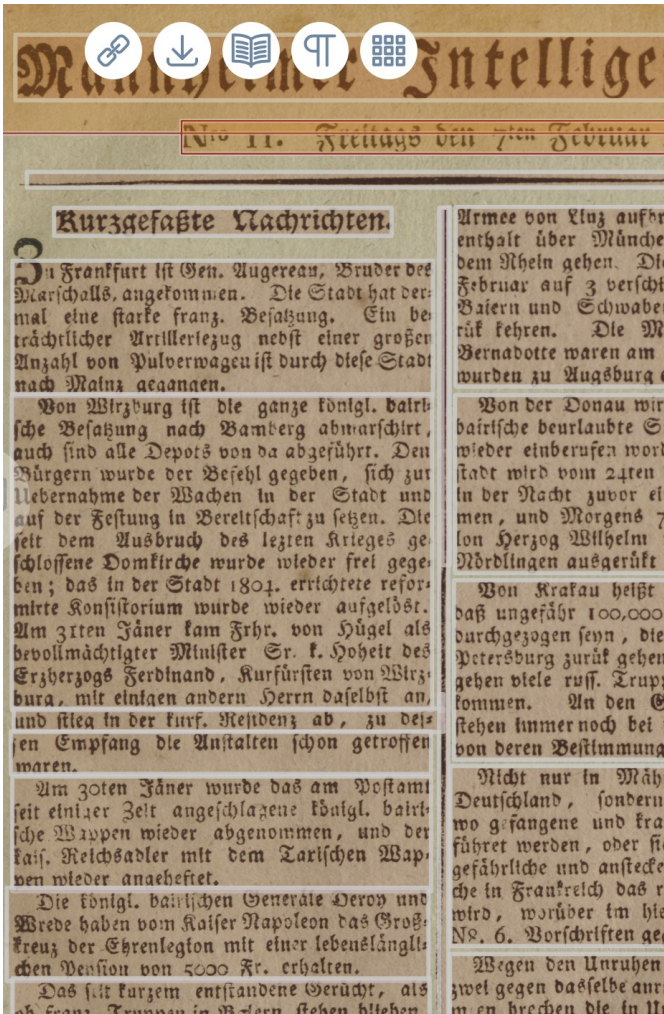
1. Land Brandenburg

Länder	Fläche in qkm	Bevölkerung	Haushalte
Brandenburg (Mecklenburg)	22,937.78	2,428	633,165
Sachsen-Anhalt	24,666.80	1,378	510,539
Thüringen	15,596.33	939	292,497
Sachsen	16,991.92	2,005	558,856

2. Sowohl. Besatzungszone insgesamt (ohne Berlin)

Länder	Fläche in qkm	Bevölkerung	Haushalte
--------	---------------	-------------	-----------

Zusammenfassung mit lokaler KI als Kitodo-Erweiterung



Mannheimer Intelligenzblatt,

Nro 1T. Freitags den 7ten Februar 1806.

KAurze gefasste Nachrichten.

Frankfurt ist Gen. Augereau, Bruder des Marfchalls, angekommen. Die Stadt hat der= mal eine ftarke franz. Befatzung. Ein be= trächtlicher Artilleriezug nebst einer großen Anzahl von Pulverwagen ist durch diese Stadt nach Malnz gegangen.

VBon Wirzburg ist die ganze könlgl. balrl= sche Befatzung nach Bamberg abmarfchlrt, auch find alle Depots von da abgeföhrt. Den Bürgern wurde der Befehl gegeben, sich zur Uebernahme der Wachen in der Stadt und auf der Festung in Bereltfchaft zu setzten. Die feit dem Ausbruch des lezten Krteges ge

llama3.2-vision:latest 1:08 PM

Der Text ist ein Auszug aus dem "Mannheimer Intelligenzblatt" vom 7. Februar 1806. Es enthält eine Sammlung von Nachrichten und Berichten aus verschiedenen Teilen Europas.

Einige der wichtigsten Punkte sind:

- General Augereau, Bruder des Marschalls, ist in Frankfurt angekommen und hat die Stadt mit einer starken französischen Besatzung übernommen.
- Die königliche bayerische Besatzung in Würzburg ist abgemarchiert und alle Depots wurden abgeführt. Die Bürger wurden aufgefordert, sich zur Übernahme der Wachen in der Stadt und auf der Festung zu setzen.
- Das reformierte Konsistorium in Würzburg wurde wieder eingerichtet und die katholische Kirche wurde wieder freigegeben.
- Der bevollmächtigte Minister des Erzherzogs Ferdinand, Kurfürst von Würzburg, Frhr. von Hügel, ist nach München gekommen.
- Die königliche bayerische Generale Derooy und Wrede haben das Großkreuz der Ehrenlegion mit einer lebenslangen Pension von 5000 Franken erhalten.
- Ein Gerücht über französische Truppen in Bayern, die bis zu den nordischen Angelegenheiten abwarten, ist ungegründet. Die Hauptquartier und der Generalstab der großen französischen Armee sollen nach München und Augsburg gehen.
- Es gibt Berichte von Unruhen in Serbien und türkischen Armeen, die gegen das Land vorrücken. Die ungarischen Truppen brechen auf, um an den Grenzen zu stehen.

Der Text enthält auch einige weitere Nachrichten über militärische Bewegungen, Krankheiten und politische Entwicklungen in

Übersetzung mit lokaler KI als Kitodo-Erweiterung

"Manheim Intelligence Newspaper,
No. 1T, Friday, February 7th, 1806.

Brief News.

General Augereau, brother of the Marshal, has arrived in Frankfurt. The city now has a strong French g artillery train, accompanied by a large number of powder wagons, has passed through this city on its v From Würzburg it is reported that the entire royal Bavarian garrison has been withdrawn to Bamberg, a removed from there as well. The inhabitants were ordered to take over the watches in the city and on 1 outbreak of the last war, the closed-off church had been reopened; the reformed consistory establish and the Catholic Church was again made accessible.

The appointed minister for the Archduke Ferdinand, Elector of Würzburg, Freiherr von Hügel, has arriv Generals Deroy and Wrede, royal Bavarian generals, have received the Grand Cross of the Legion of H pension of 5,000 francs.

A rumour that French troops are stationed in Bavaria until the northern affairs are settled is unfounded. The main headquarters and general staff of the large French army will depart for Munich and Augsburg.

From Krakow it is reported that about 10,000 Russian troops have passed through, who will return to Petersburg via Therespol. Many Russian troops also march through Lemberg from Moravia. There are still around 150,000 Russian soldiers at the borders of Galicia.

"Le Mannheimer Intelligenzblatt,
No 1T, Vendredi 7 février 1806.

Nouvelles brèves.

Général Augereau, frère du maréchal, est arrivé à Francfort. La ville est désormais sous la garde d'une forte garnison française. Un train considérable d'artillerie, accompagné de nombreux wagons de poudre, a passé par cette ville en route pour Mayence.

De Wurtzburg, on rapporte que toute la garnison bavaroise royale s'est retirée à Bamberg et que tous les dépôts ont été transférés. Les habitants ont été chargés de prendre les tours de guet dans la ville et sur la forteresse. Depuis le début de la dernière guerre, l'église fermée avait été rouvertes ; le consistoire réformé établi en 1804 a été réinstauré et l'église catholique est à nouveau accessible.

Le ministre délégué pour l'archiduc Ferdinand, électeur de Wurtzburg, baron von Hügel, est arrivé à Munich.

Les généraux Deroy et Wrede, généraux bavarois royaux, ont reçu la Grand-Croix de la Légion d'Honneur avec une pension vitale de 5 000 francs.

Un bruit selon lequel des troupes françaises sont stationnées en Bavière jusqu'à ce que les affaires du Nord soient réglées est infondé. Le quartier général et l'état-major de la grande armée française se dirigent vers Munich et Augsburg.

De Cracovie, on rapporte qu'environ 10 000 troupes russes ont passé par là pour retourner à Saint-Petersbourg via Therespol.

De nombreuses troupes russes marchent également à partir de Moravie par Lemberg. Il y a encore environ 150 000 soldats russes aux frontières de Galicie.

Übersetzung von deutschen, lateinischen, englischen, französischen, italienischen, spanischen Texten in die gleichen Zielsprachen und weitere (chinesisch, arabisch, ...)

Aktualisierung auf neueste Releases von Kitodo.Production u. Kitodo.Presentation

- Upgrade für Kitodo.Presentation und Kitodo.Production für UB Mannheim schon lange geplant (und inzwischen längst überfällig).
- Gewünschte Voraussetzungen für Kitodo.Presentation:
 - Unterstütztes TYPO3 Release (momentan TYPO3 12) (✓)
- Gewünschte Voraussetzungen für Kitodo.Production:
 - OpenJDK 17 ✓
 - OpenSearch ✓

„Upgrade“ für Zeitungs-PDF

- Häufig findet man u. a. in Stadtarchiven PDF-Dateien mit historischen Zeitungen. Dies entspricht nicht den Praxisregeln „Digitalisierung“ der DFG.
- Ziel: PDF-Dateien mit gescannten Zeitungen aufbereiten für die Präsentation im DFG-Viewer und im Deutschen Zeitungsportal. Notwendige Schritte:
 - Bilddateien der eingescannten Seiten möglichst verlustfrei extrahieren (inkl. Drehen und Beschneiden, sofern erforderlich) ✓
 - Bookmarks extrahieren für weitere Verwendung in den Metadaten ✓
 - Texterkennung mit Erzeugung von ALTO XML für jede Seite ✓
 - METS/MODS erzeugen x (gesucht: vorkonfiguriertes Kitodo.Production)
 - Daten online zugänglich machen

- Validierungsskript für METS-Dateien (Zeitungen und andere Medien)
→ <https://github.com/UB-Mannheim/Reichsanzeiger/tree/main/scripts/validate>
- Schematron-Validierungen der Fachstelle Bibliothek der Deutschen Digitalen Bibliothek
→ <https://github.com/Deutsche-Digitale-Bibliothek/ddb-metadata-schematron-validation>
- eScriptorium der Universitätsbibliothek Mannheim
→ <https://ocr-bw.bib.uni-mannheim.de/escriptorium/>