



zeit.punktNRW – ein Zwischenstand zur Mitte der dritten Förderphase

Dr. Michael Herkenhoff (ULB Bonn)

Visual Library-Anwendentreffen 2024 in Mannheim
11. März 2024

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen



UNIVERSITÄTS- UND LANDESBIBLIOTHEK





Vorträge 2019 und 2024

- zeit.punktNRW – ein Zwischenstand (Bonn, 28.03.2019)
- zeit.punktNRW – ein Zwischenstand zur Mitte der dritten Förderphase (Mannheim, 11.03.2024)
- Fragen:
 - Was hat sich gegenüber der ursprünglichen Projektplanung geändert?
 - Wie ist der momentane Projektstand?



Inhaltliche Festlegungen 2016

- Digitalisierung von Mikroformen (Mikrofilm, Mikrofiche)
- Übernahme bereits vorhandener Digitalisate
- Lokale, Kreis- und Regionalzeitungen allgemeinen Inhalts
- Orientierung an den Grenzen des heutigen Bundeslandes Nordrhein-Westfalen
- Zeitraum 1801 – 1945
- technische Features: Kalendersicht, „Visualisierungskomponente“, Download von Zeitungsausgaben als PDF-Dateien, keine OCR-Erkennung



Inhaltlicher Stand 2024

- Digitalisierung von Mikroformen (Mikrofilm, Mikrofiche) und **Direktdigitalisierung**
- Übernahme bereits vorhandener Digitalisate
- Lokale, Kreis- und Regionalzeitungen allgemeinen Inhalts
- Orientierung an den Grenzen des heutigen Bundeslandes Nordrhein-Westfalen
- Zeitraum 1801 – 1945
- technische Features: Kalendersicht, „Visualisierungskomponente“, Download von Zeitungsausgaben als PDF-, **Volltext- und alto.xml** Dateien, ~~keine~~ OCR-Erkennung



Mengengerüst: Kalkulation 2016

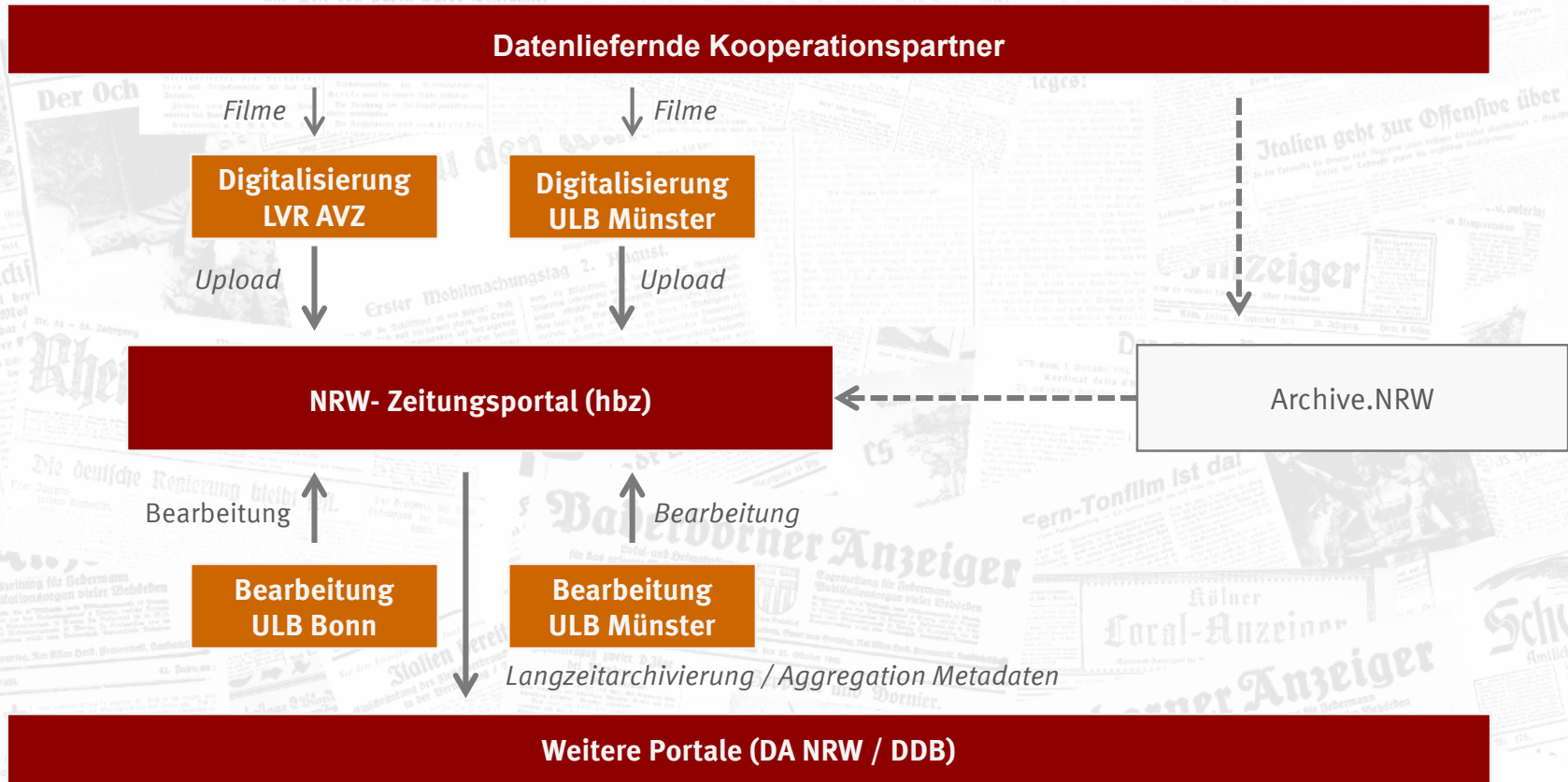
- Digitalisierung von 20.000.000 Zeitungsseiten innerhalb von 10 Jahren
- 1. Förderphase (2017 – 2019): Digitalisierung von 4.800 Masterfilmen mit ca. 6.000.000 Zeitungsseiten
- 1. Förderphase (2017 – 2019): Migration von weiteren 2.000.000 Zeitungsseiten (bereits vorhandene Digitalisate) in das NRW-Portal
- Kalkulierter Speicherbedarf für 20.000.000 Seiten:
 - 500 TB Plattenplatz für die Scans
 - 150 Terabyte Plattenplatz für die Präsentation



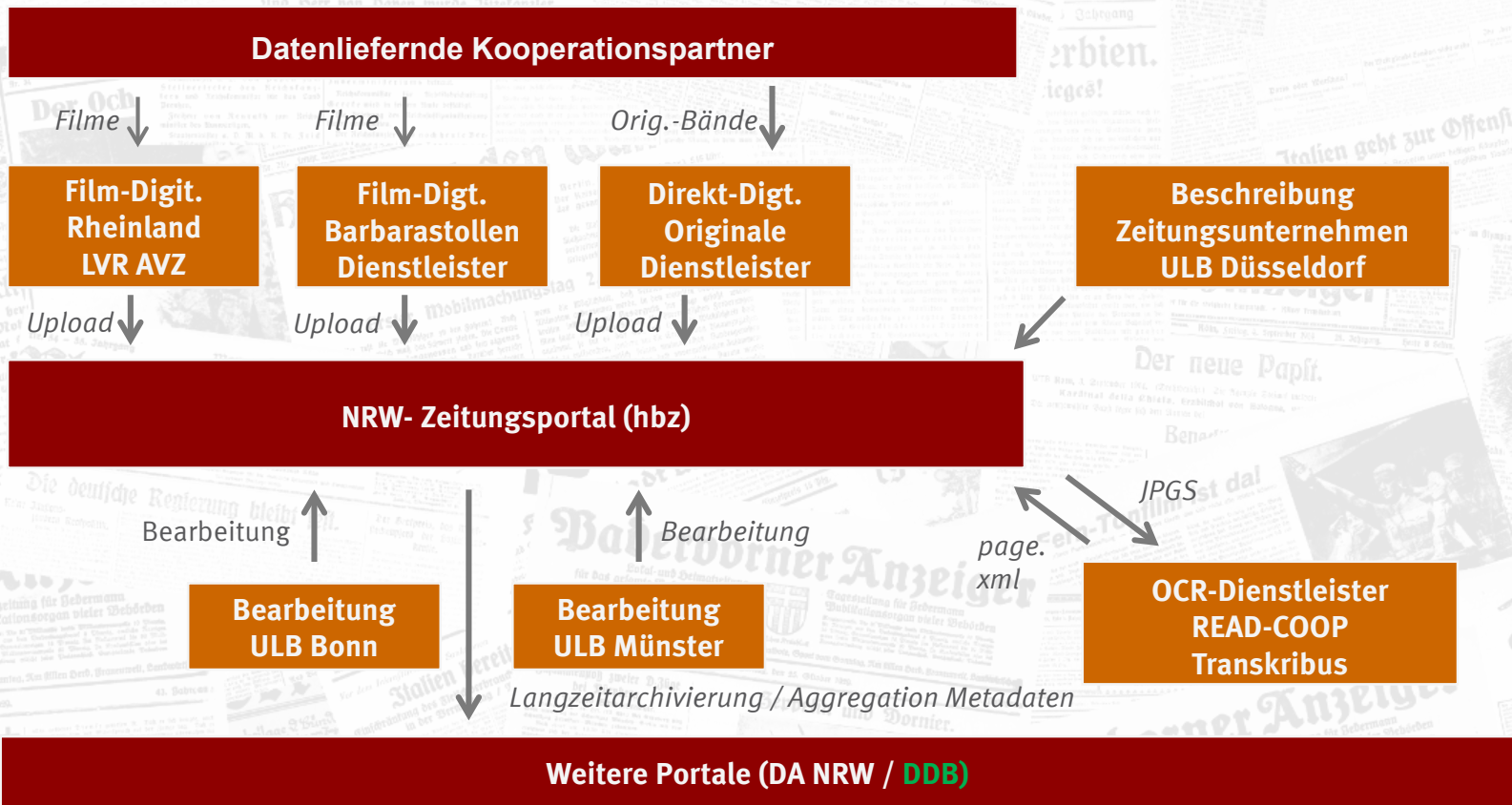
Mengengerüst: Annahme 2024

- Digitalisierung von **27.000.000** Zeitungsseiten innerhalb von **12 (?)** Jahren (Projektdauer 2017 – 2028)
- Kalkulierter Speicherbedarf für **27.000.000** Seiten:
 - **600** TB Plattenplatz für die Scans
 - 150 TB Plattenplatz für die Präsentation
 - **24 TB SSD-Speicher für die Volltexte**

Workflow 2019



Workflow 2024





OCR-Einführung

- 1. Förderphase (2017 – 2019): keine OCR-Erkennung eingeplant
- 2. Förderphase (2020 – 2022): Geld für OCR-Erkennung in das Projekt eingestellt
- Ausschreibung Frühjahr 2020
 - Wie viele Seiten können für eine bestimmte Summe durch eine OCR-Erkennung laufen?
 - Dienstleister erhielten 10 Samples zur Texterkennung



Abgegebene Angebote

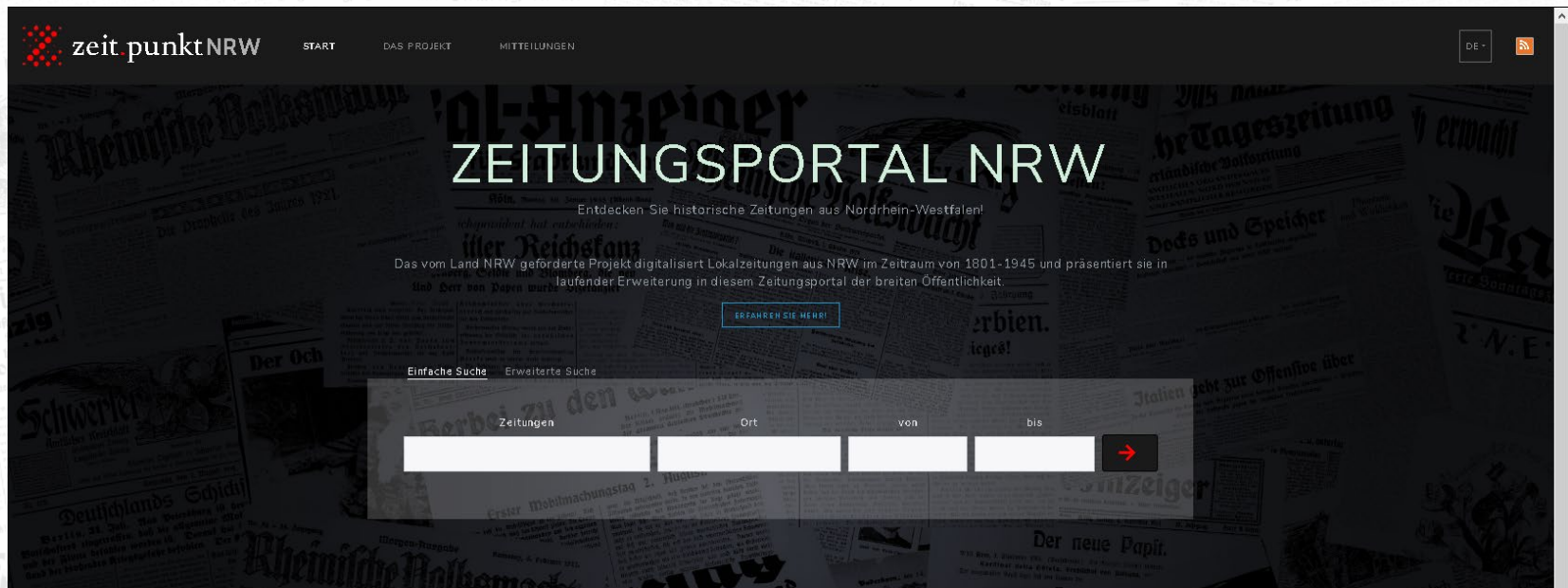
- Vorgegebener Sollwert: 95% Zeichengenauigkeit
- Überprüfung der Samples im Mai 2020 (3 SHKs, 17 Tage)
 1. Angebot: ABBYY-Finereader, durchschnittlicher Erkennungsrate deutlich unter 95% (Samples 5–7: 87,6% – 92,8%)
 2. Angebot: Tesseract: durchschnittliche Erkennungsrate: 95,26%
 3. Angebot: Transkribus: durchschnittliche Erkennungsrate: 98,51 %
- Ergebnis bestätigt durch 2. Ausschreibung zur „OCR-Erkennung“ der „Kölnische Zeitung“ (DFG-Projekt)
- Vergabe jeweils an READ COOP



OCR-Durchgänge

- „Kölnische Zeitung“ Juli 2021
- zehn benutzungstärkste Zeitungen August bis November 2021
 1. OCR-Durchgang A-Z Dezember 2021- Dezember 2022 (Stichtag 20.09.2021)
 2. OCR-Durchgang ab Januar – Juni 2023
 3. OCR-Durchgang seit Juli 2023
 4. OCR-Durchgang 2024; dann 15. Millionen Seiten texterkannt

NRW-Zeitungsportal: Frontend 2018



FINDPUNKTE

- ZEITUNGEN
- KARTE
- ORTE
- KALENDER



Änderung Frontend: Herbst 2021



Facettierung Suchergebnisse

927533 TREFFER FÜR VOLLTEXT = BISMARCK

ALLE DETAILS

Treffer pro Seite: Reihenfolge: Blättern:



GLADBACHER MERKUR 8 (31.10.196) 129

← 2 Treffer im Volltext

M.Gladbach : Fongern & Zester, 1. Jahrgang, Nr. 1 (14. September 1889)-16. Jahrgang, No. 194 (31. Dezember 1904)

+ Details



LE NOUVELLISTE POLITIQUE D'ALLEMAGNE (18.2.1783) 28

← 1 Treffer im Volltext

Cologne, 1780 - 1787[?]

+ Details



ALLGEMEINE POLITISCHE NACHRICHTEN (6.9.1807) 72

← 1 Treffer im Volltext

Essen : Expedition, 1800,1(2.Jan.)-53(3.Juli)

+ Details



JOURNAL DES NIEDER- UND MITTEL-RHEINS (30.7.1814) 20

← 1 Treffer im Volltext

Aachen : Beaufort, 1.1814,1(15.März) - 8.1816,48(20.Apr.); damit Ersch. eingest.

+ Details



KÖLNISCHE ZEITUNG (7.3.1815) 38

← 1 Treffer im Volltext

Köln : DuMont Schauberg, Nro. 50 (1. Janner 1803)-Nro. 182 (22. Septemb.); 1803, Nro. 8 (8. Oktober 1803)-1945, Nr. 25 (31. Januar 1945)

+ Details

ZEITUNGEN

Kölnische Zeitung, 1803-1945	24875
Dortmunder Zeitung, 1874-1939	21180
Düsseldorfer Zeitung, 1814-1848	19231
Rhein- und Ruhrzeitung, 1852-1940	18354
Echo der Gegenwart, 1848-1935	17989
Hagener Zeitung, 1864-1945	14876
Aachener Anzeiger, 1885-1943	14389
Essener Volks-Zeitung, 1869-1941	13851
Remscheider Zeitung: Lokal-Anzeiger, 1869-1923	13747
Remscheider General-Anzeiger, 1890-1945	13074

ZEIGE 40 WEITERE

ORTE

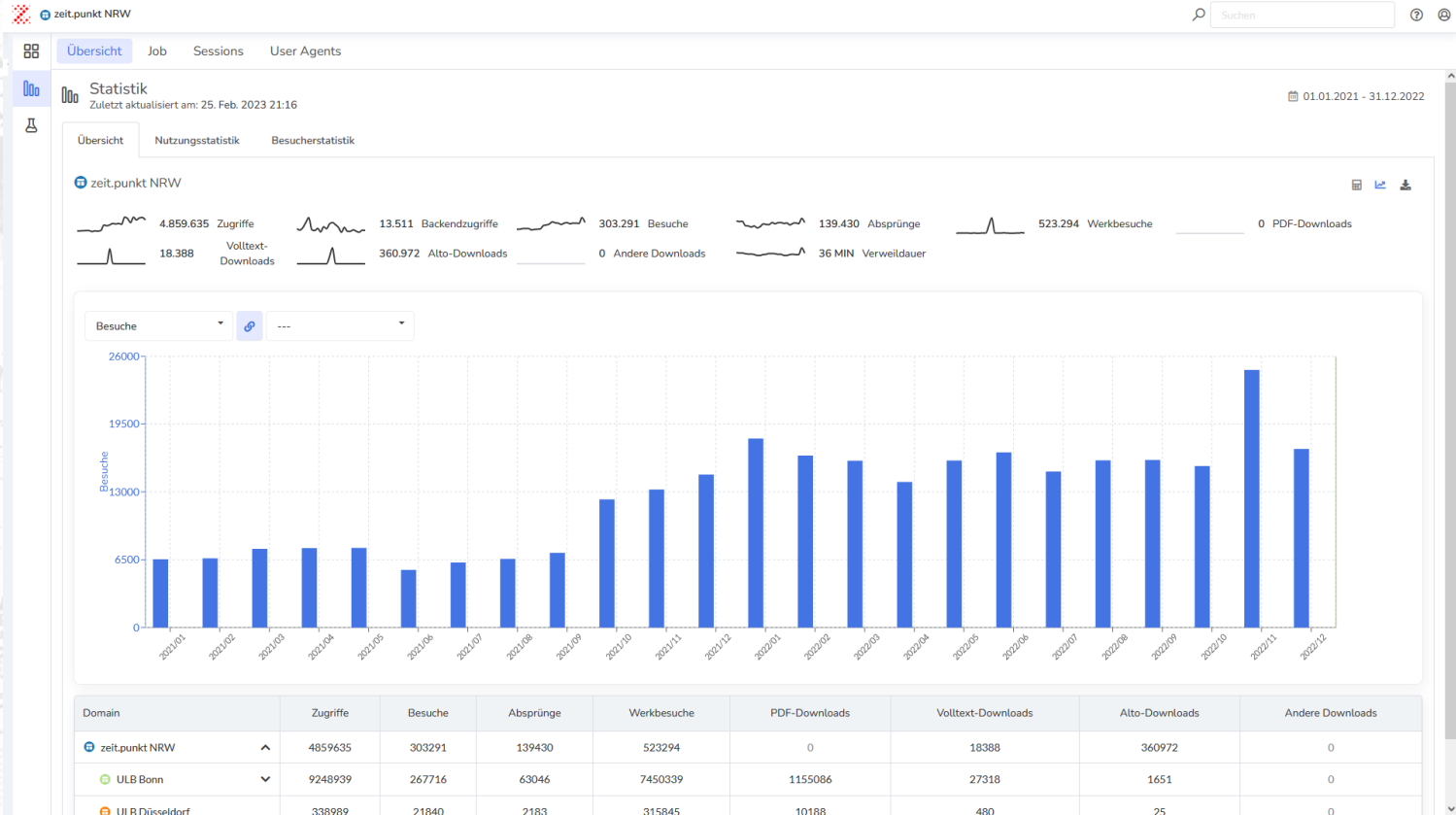
Köln	57108
Dortmund	55679
Bielefeld	54389
Solingen	46137
Düsseldorf	43038
Essen	42794
Bonn	40365
Aachen	38448
Duisburg	36576
Gelsenkirchen	35499

ZEIGE 40 WEITERE

ZEITRÄUME

191-200	1
1781-1790	1
1801-1810	1
1811-1820	4
1821-1830	12
1831-1840	14
1841-1850	64

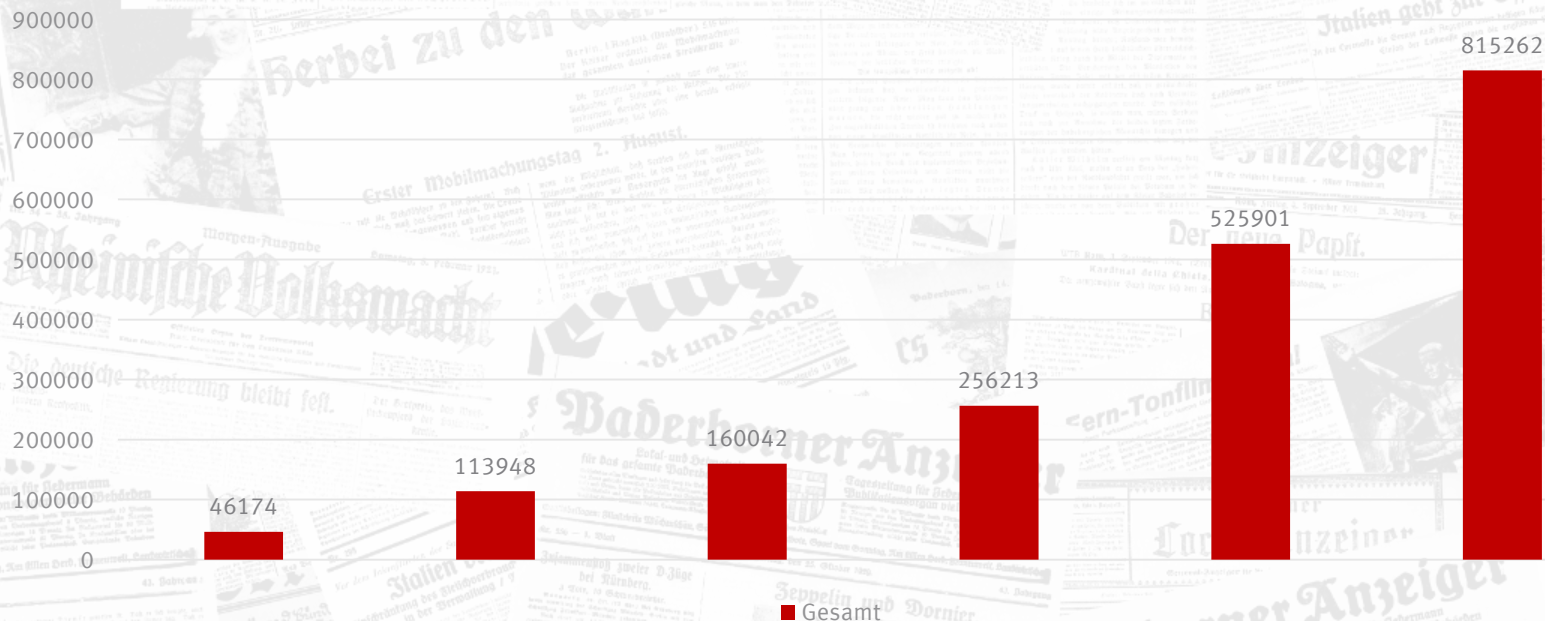
Zugriffsstatistik 2021/22 (Monate)





Zugriffsstatistik 2018-2023

Besuche: 2018-2023





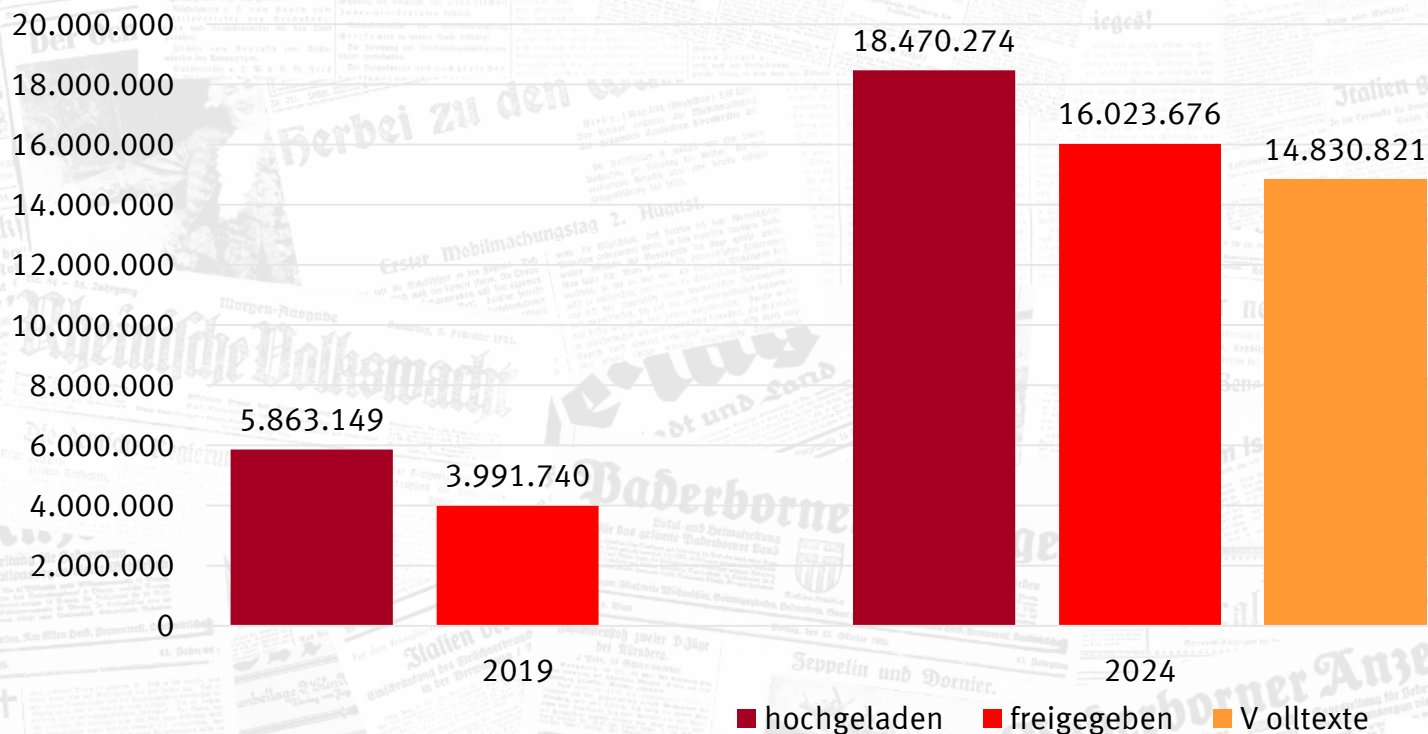
Partner (Stand: 09.03.2024)

- Projektpartner: 4 (5)
- Kooperationspartner: 108
- Archive/Forschung 92
- Bibliotheken 6
- Verlage 4
- Vereine 4
- Firmen 1
- Museen 1



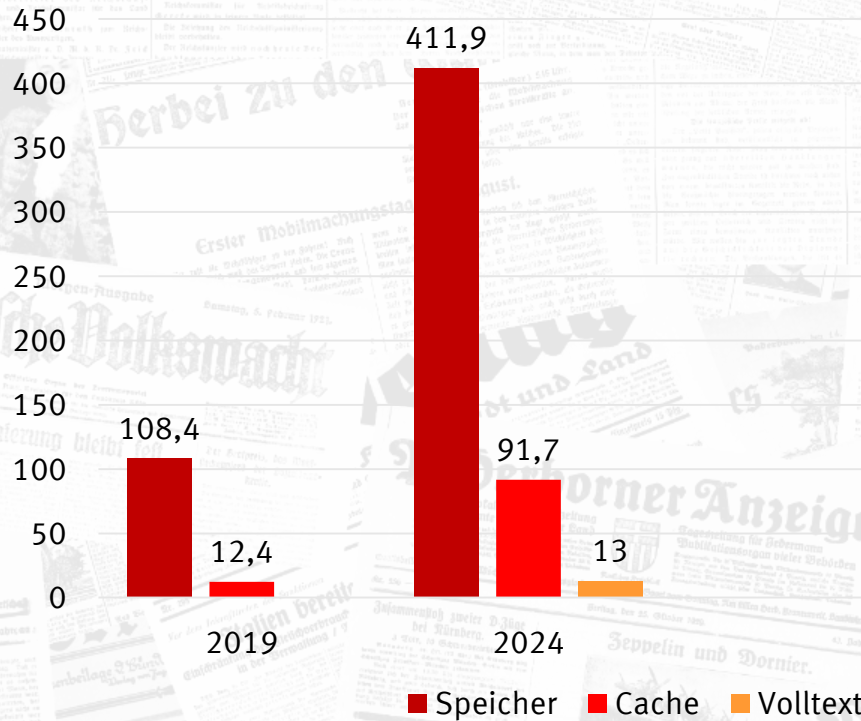
Images 2019 versus 2024

Images (Stand: 27.03.2019 u. 09.03.2024)



Speicher 2017 versus 2019

Speicherplatz (Angaben in TB)





Verbleibendes Mengengerüst Anfang 2024

- digitalisiert Ende 2023:	18.210.000
- Noch zu digitalisieren:	
- Filmdigitalisierung Rheinland	3.450.000
- Barbarastollen	2.500.000
- Direktdigitalisierung Westfalen	1.830.000
- Direktdigitalisierung Rheinland	1.200.000
- Summe	27.190.000

Facebookgruppe: Zeitungsportal NRW

The screenshot shows the Facebook interface for the group 'Zeitungsportal NRW'. The group has 2,990 members and is public. The cover photo features the text 'ZEITUNGSPORTAL NRW' and 'Entdecken Sie historische Zeitungen aus Nordrhein-Westfalen!'. Below the cover is a search bar with fields for 'Zeitung', 'Ort', 'von', and 'bis'. The left sidebar contains navigation options like 'Community-Startseite', 'Übersicht', and 'Admin-Tools'. The main content area shows the group name, member avatars, and a 'Diskussion' tab. The bottom right corner includes an 'Info' section with details about the group's purpose and privacy settings.

Danke für Ihre Aufmerksamkeit!

zeit.punktNRW

Ministerium für
Kultur und Wissenschaft
des Landes Nordrhein-Westfalen



UNIVERSITÄTS-
UND LANDESBIBLIOTHEK



Dr. Michael Herkenhoff (ULB Bonn) 11.03.2024