Erschließung, Präsentation, Zitation: Herausforderungen der Zeitungsdigitalisierung

13. scantoweb-Workshop
27. Februar 2014

Universitätsbibliothek Johann Christian Senckenberg

1. Jahrgang. Frankfurter Universitäts-Zeitung Nummer 8.

Sommer 1914 mit amtlichen Mitteilungen. 8. Juli 1914.

Inhalt der Nr. 8: Eine Zeppelinfahrt. — Darmstädter Künstler-Ausstellungen 1914. — Rechts- und Wirtschaftsphilosophie III: Volksschädliche Wirkungen unseres technisch vollkommenen Hypothekenwesens. Die kommunalpolitischen Aufgaben der Wohnungsreform. — Bismarckfeier der Deutschen Studentenschaft 1915. — Zur Frage der Reformbedürftigkeit des preußischen Studentenrechts. — Bücher und Zeitschriftenschau.

No. 9 erscheint am 17. Juli.

Eine Zeppelinfahrt.

Von Ernst Fischer.

In voller Fahrt.

Kay Heiligenhaus
semantics GmbH, Aachen

s‹e›mantics
Kommunikationsmanagement GmbH

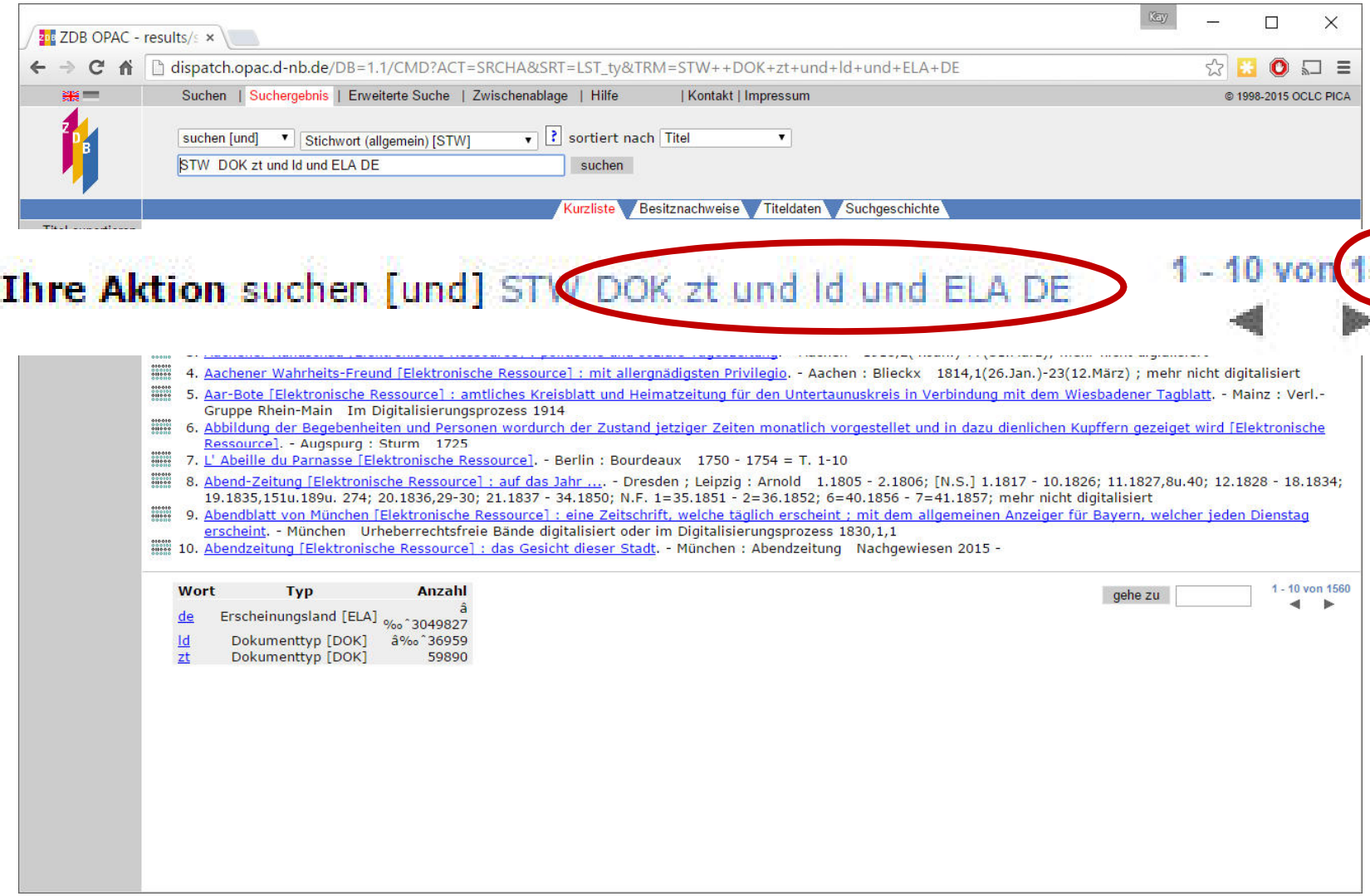
I.

Ausgangslage und Rahmenbedingungen

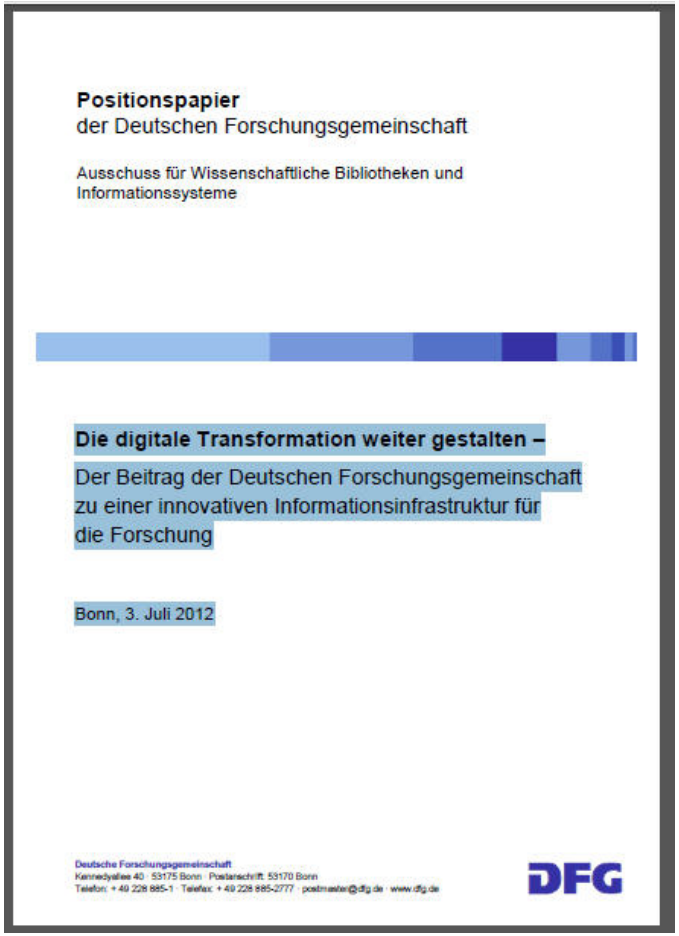
Ausgangslage: Stellenwert, Erschließung, Zugang

- Herausragender Stellenwert der historischen Zeitungsbestände für alle Bereiche der Kulturwissenschaften
- Heterogene Überlieferungslage, prekäre Erhaltungszustände, oft ungenügende Erschließungssituation, eingeschränkte Zugangsmöglichkeiten
- Quantitative Herausforderung
  - → Zeitungen in der ZDB ~ 50.000 Titel
  - → davon in Deutschland ~ 30.000 Titel
  - → davon vor 1945 ~ 16.000 Titel
- Desiderate
  - → Nachweis auf Ausgabenebene
  - → Überregionale Recherchierbarkeit
  - → Standortunabhängige Verfügbarkeit
  - → Nutzung einheitlicher Standards

Ausgangslage: Digitalisierung



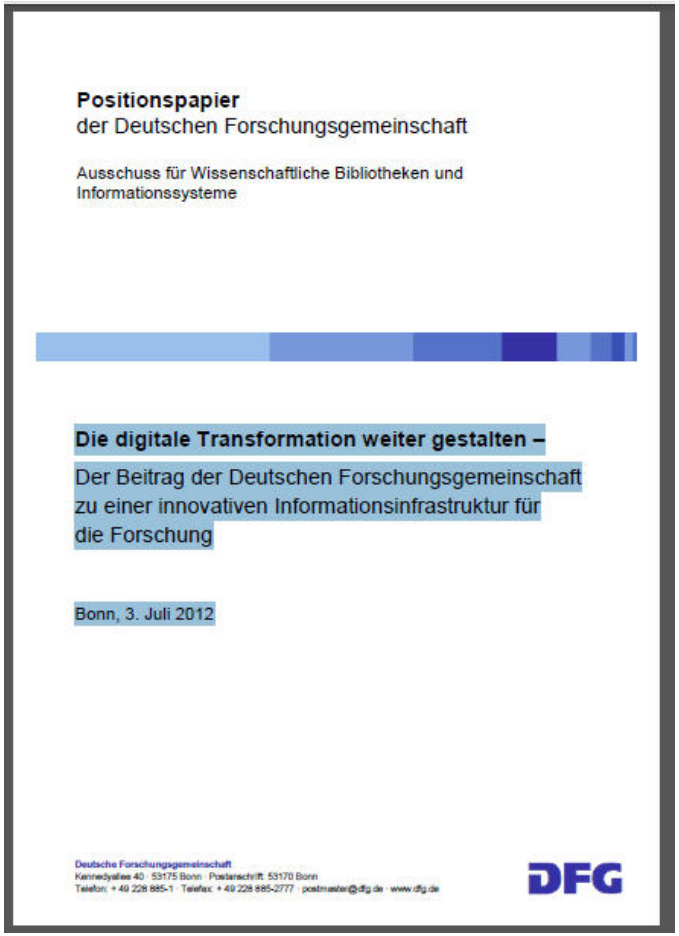
DFG-Positionspapier (2012): Digitalisierung



Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme: *Die digitale Transformation weiter gestalten* – **Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung**, *Bonn, 3. Juli 2012.*

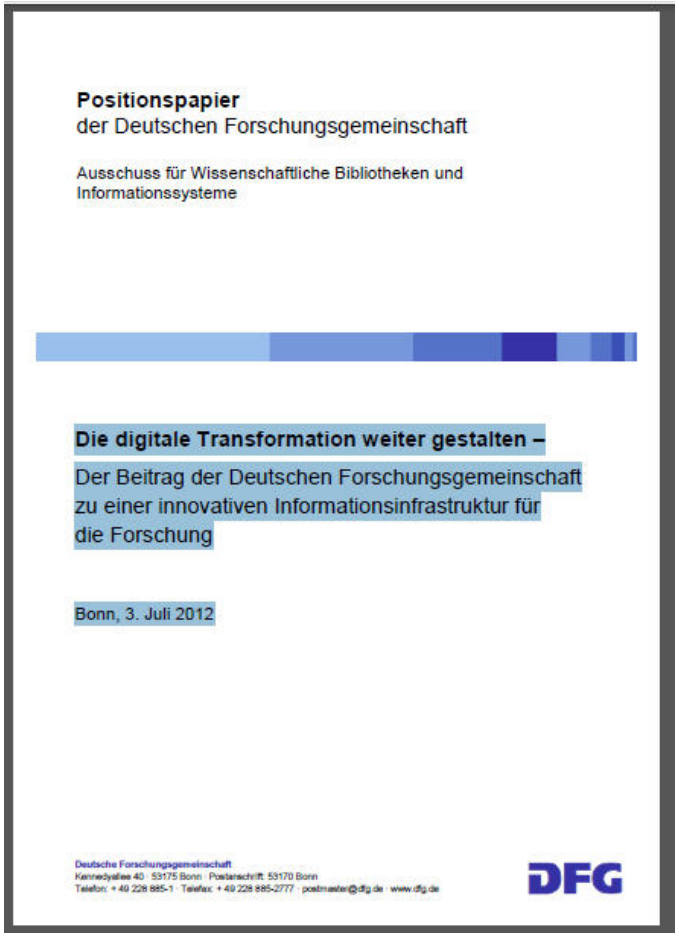
http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf

DFG-Positionspapier (2012): Förderschwerpunkte



Zur weiteren Umsetzung der digitalen Transformation analoger Überlieferung wird sich der **Schwerpunkt der Förderaktivitäten** im Bereich Digitalisierung nach und nach von **Drucken** hin zu **historischen Zeitungen** und **unikalem Material** […] verschieben. (S. 10)

DFG-Positionspapier (2012): Pilotprojekt Zeitungen



Positionspapier
der Deutschen Forschungsgemeinschaft

Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme

Die digitale Transformation weiter gestalten –
Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung

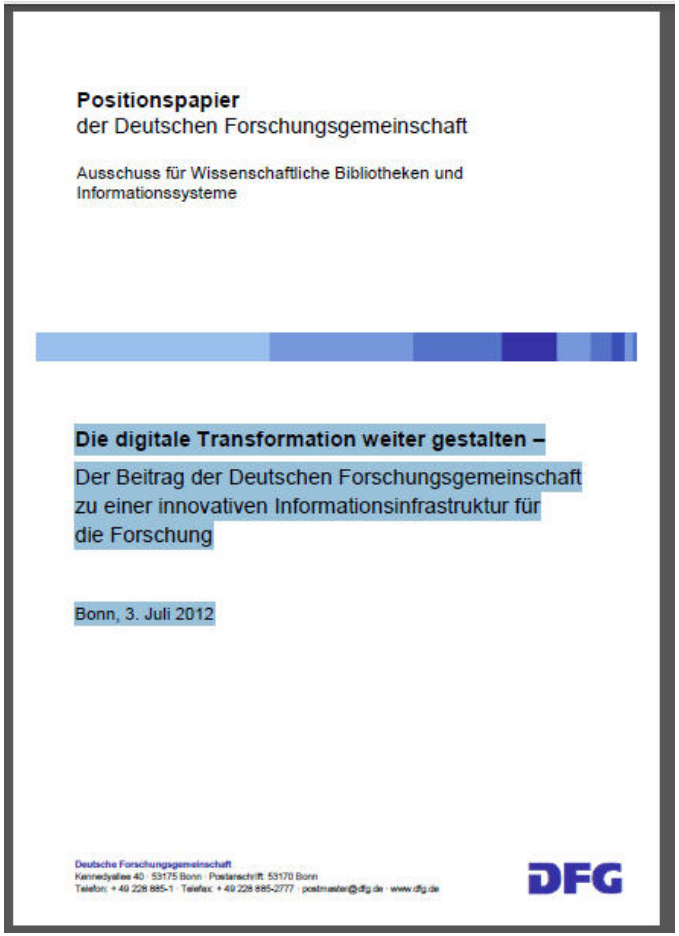
Bonn, 3. Juli 2012

Deutsche Forschungsgemeinschaft
Kennedyallee 40 · 53175 Bonn · Postanschrift: 53170 Bonn
Telefon: + 49 228 885-1 · Telefax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

DFG

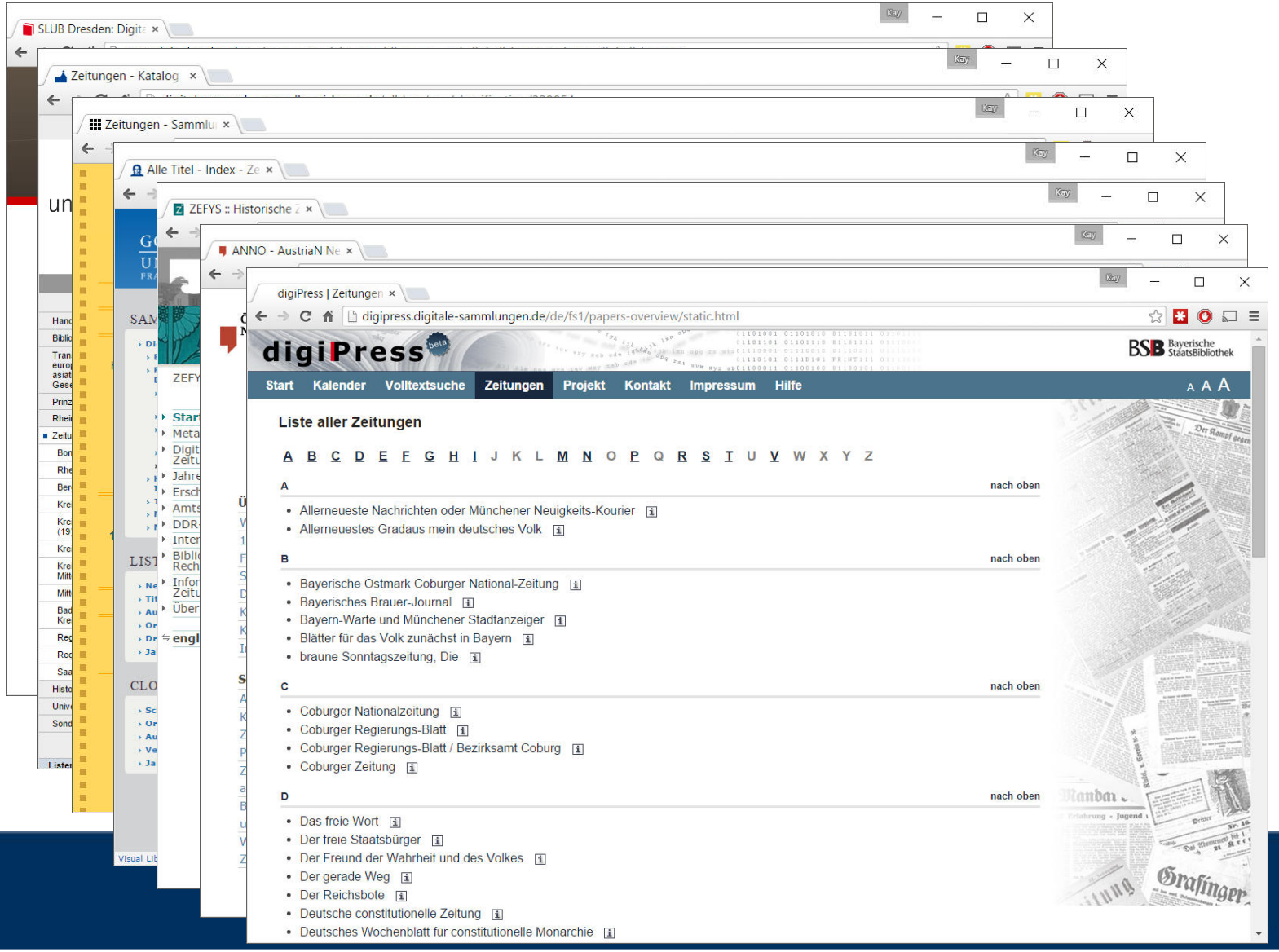
Bei der […] **Förderung von Digitalisierungsmaßnahmen im Bereich historischer Zeitungen** […] wird es in erster Linie darum gehen, im Rahmen von **Pilotphasen methodische, organisatorische und technische Rahmen-bedingungen** […] zu definieren. (S. 10)

DFG-Positionspapier (2012): Automatisierung



Vor dem Hintergrund technischer Weiterentwicklungen sollte […] eruiert werden, wie **automatisierte Verfahren** zur **vertieften Erschließung** […] Anwendung finden können. Dabei soll auch die Entwicklung neuer Methoden, zum Beispiel im Bereich des **Machine Learning**, gefördert werden. (S. 12)

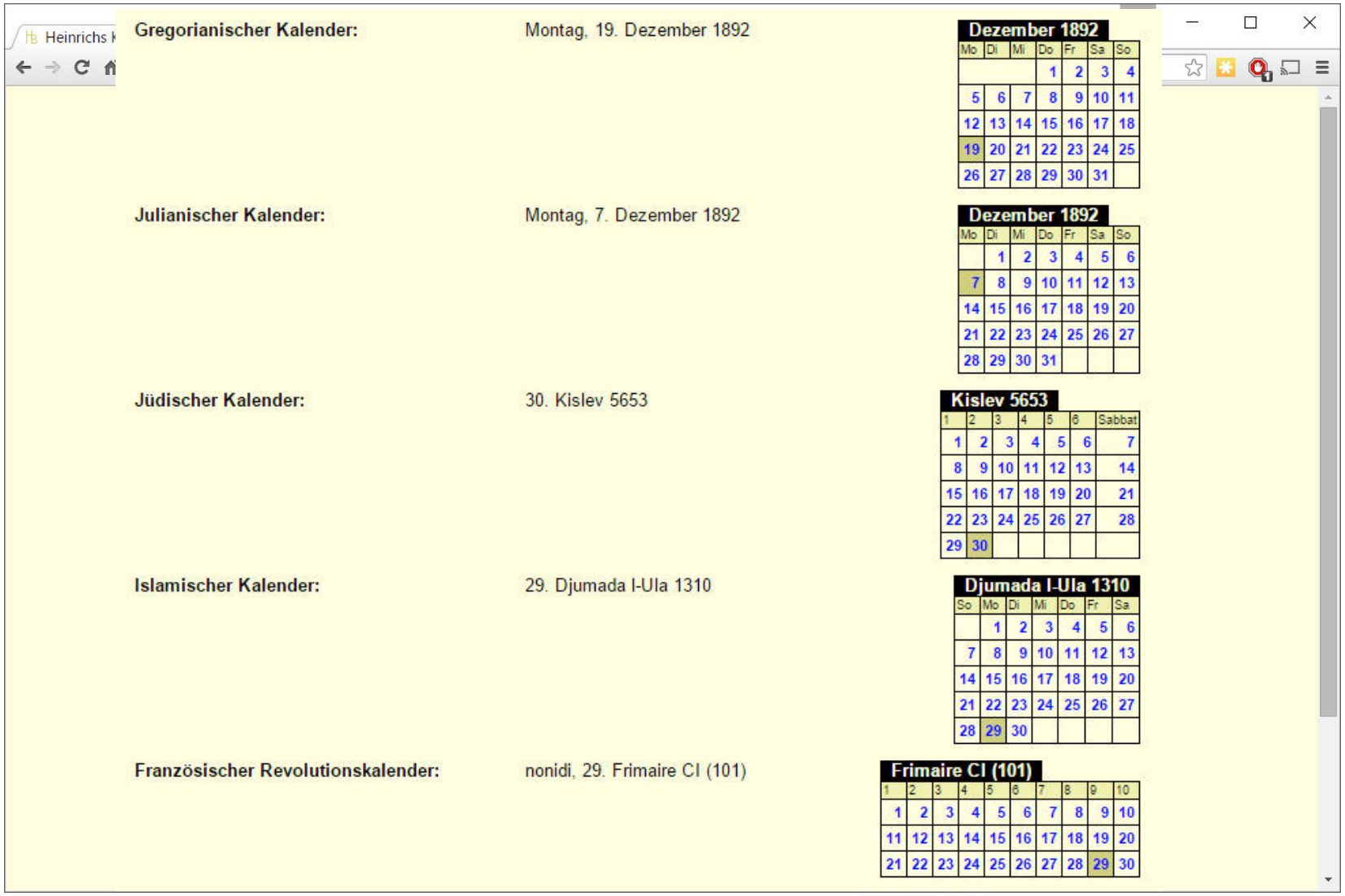
Entwicklungen in den zurückliegenden Jahren im deutschen Sprachraum



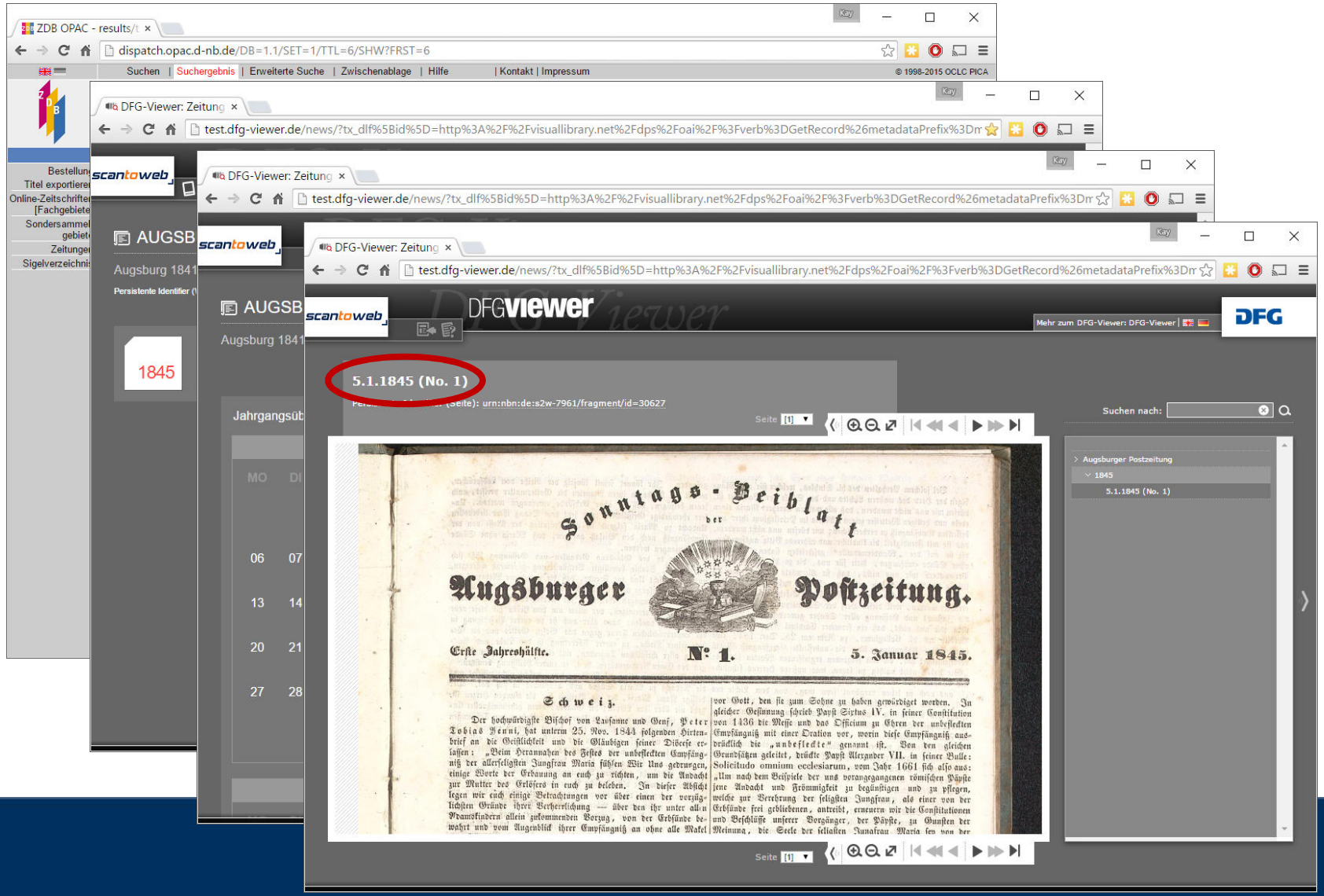
II.

Mindestanforderungen an die Erschließung

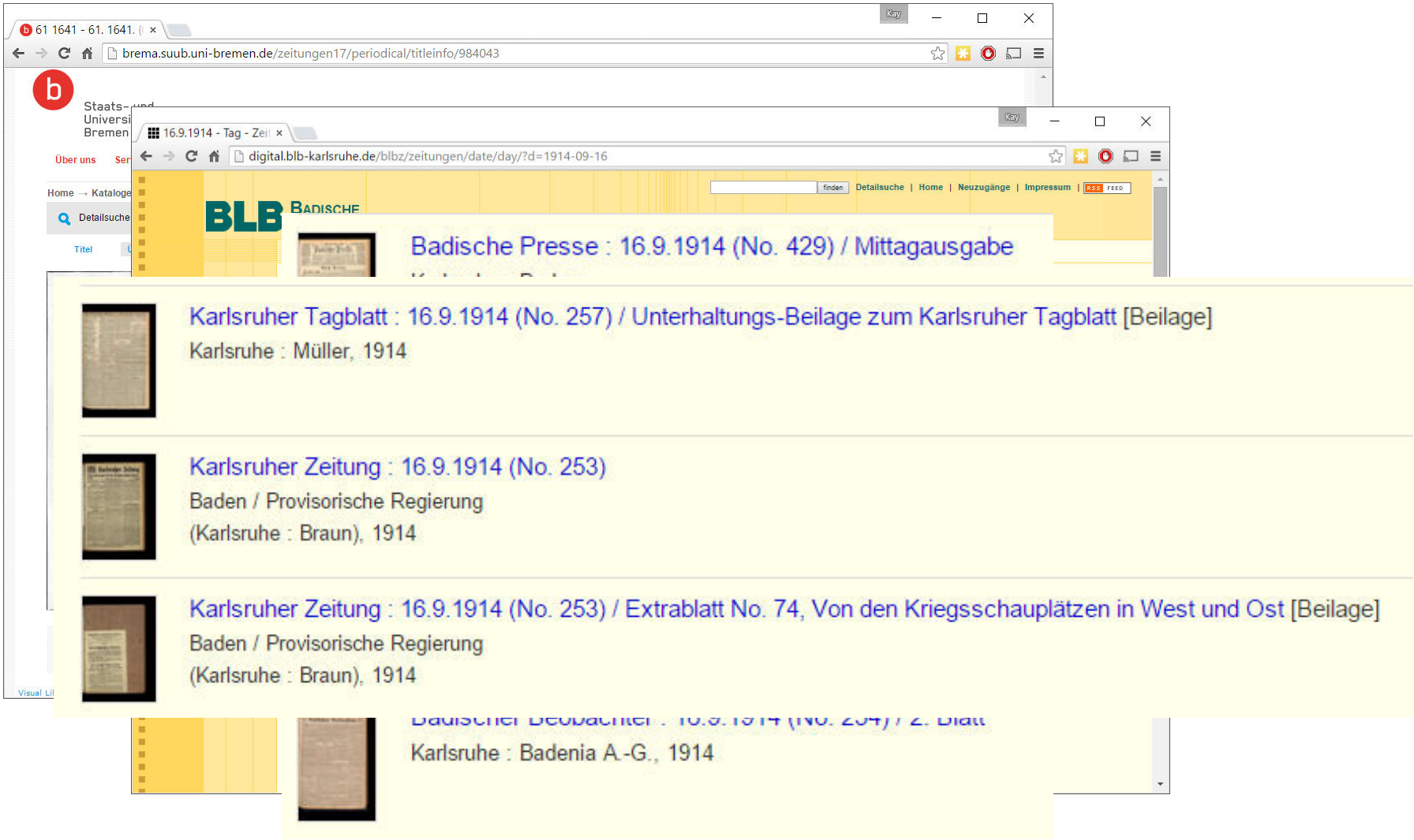
Einführung


Heinrichs K
Gregorianischer Kalender:
Montag, 19. Dezember 1892
Julianischer Kalender:
Montag, 7. Dezember 1892
Jüdischer Kalender:
30. Kislev 5653
Islamischer Kalender:
29. Djumada I-Ula 1310
Französischer Revolutionskalender:
nonidi, 29. Frimaire CI (101)
Dezember 1892
Mo Di Mi Do Fr Sa So
1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 31
Dezember 1892
Mo Di Mi Do Fr Sa So
1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31
Kislev 5653
1 2 3 4 5 6 Sabbat
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30
Djumada I-Ula 1310
So Mo Di Mi Do Fr Sa
1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30
Frimaire CI (101)
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30

Mindestanforderung: Von der ZDB zu einer strukturierten Navigation im DFG-Viewer



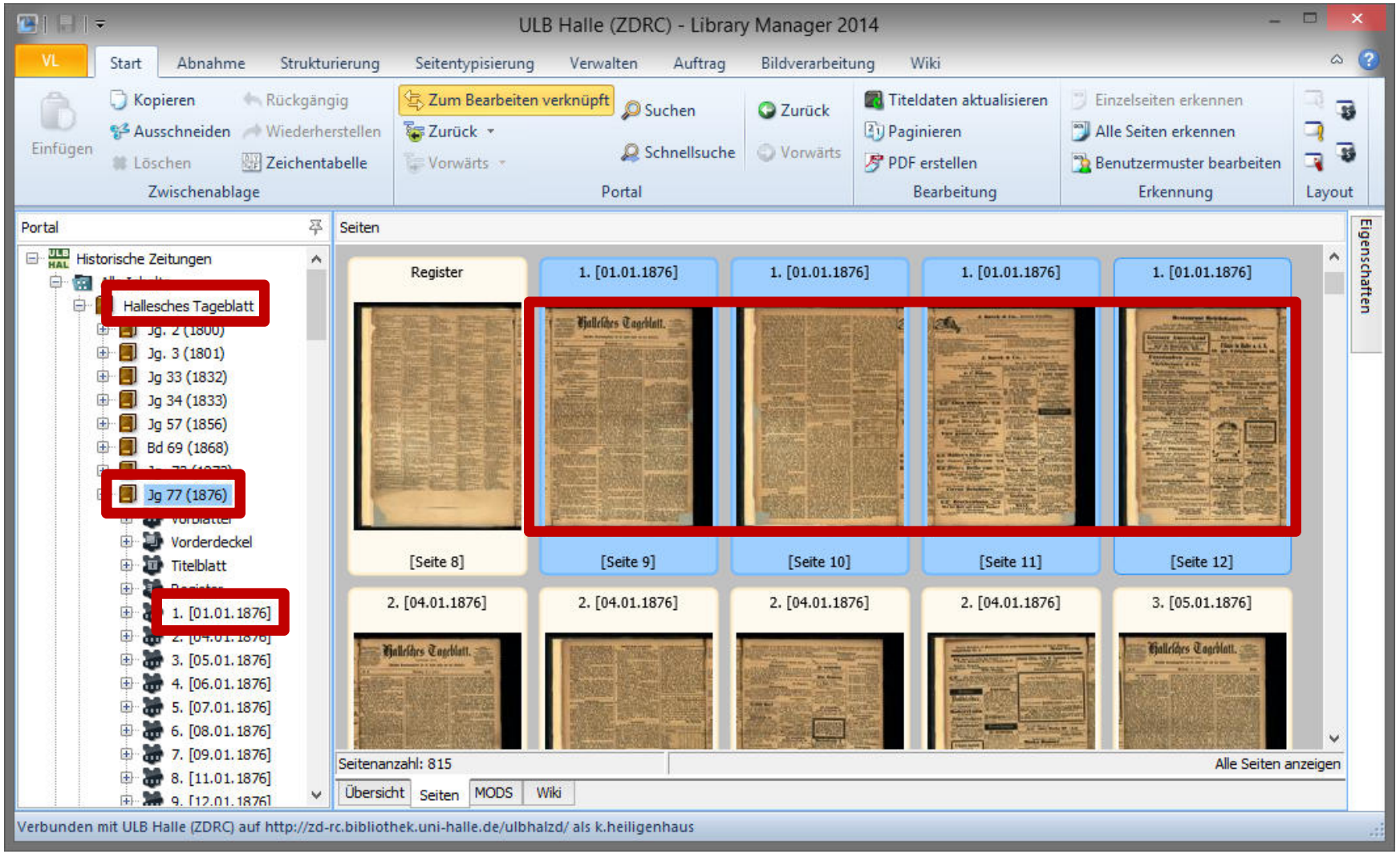
Mindestanforderung: Besonderheiten



III.

Erschließung in der Praxis

Ist-Stand: Manuelle Strukturdatenerfassung



Mengengerüst und Kosten: Minimal-Erschließung

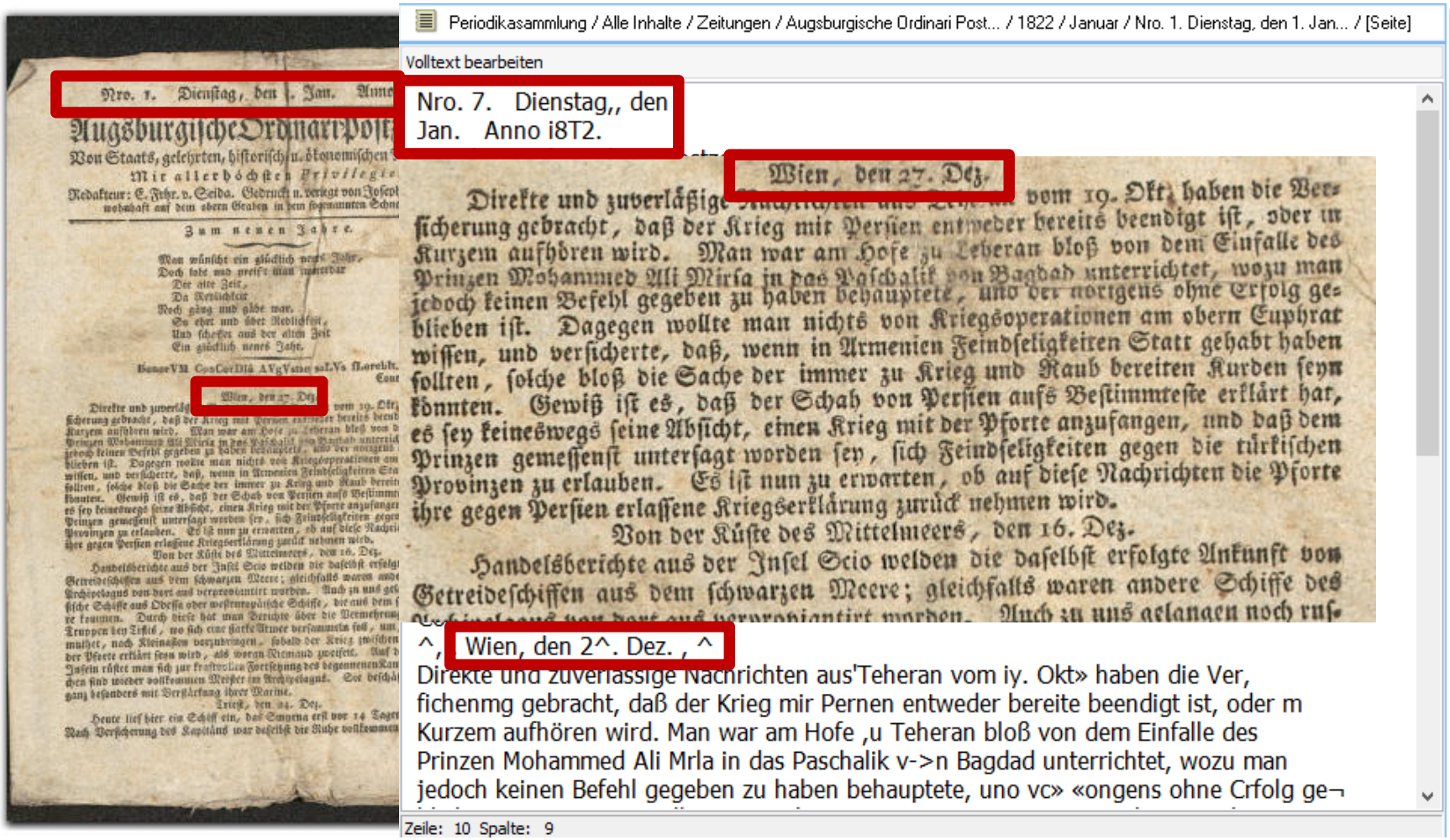
- **Umfang**: ~ **160 Mio. Strukturdaten** (~ 16.000 Titel, ~ 1 Mrd. Seiten)
- **Aufwand**: ~ **1.000 Personenjahre** (80 Strukturdaten/Stunde, 8 Stunden/Tag, 250 Arbeitstage/Jahr)
- **Kosten: ~ 35 Mio. EUR** (~ 2.161 EUR/pro Titel, ~ 0,035 EUR/pro Seite)

➢ **Herausforderung:** Reduktion der Erschließungskosten.

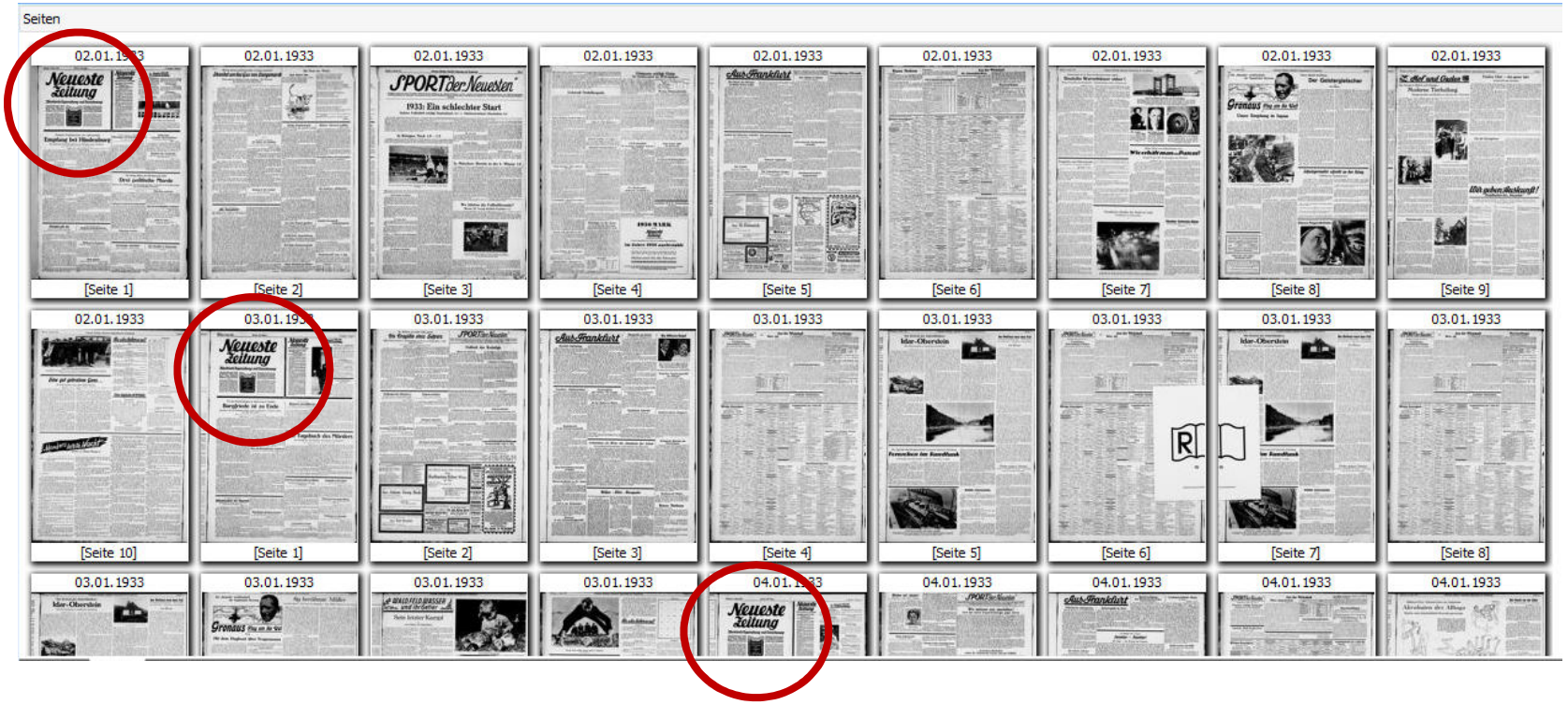
Lösungsoption I: Nutzung der Layout-Analyse und Text-Erkennung von Abbyy-Finereader

```
<alto xsi:schemaLocation="http://schema.ccs-gmbh.com/ALTO http://www.loc.gov/ndnp/alto_1-1-041.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://schema.ccs-gmbh.com/ALTO">
  - <Layout>
    - <Page WIDTH="182" HEIGHT="279" PHYSICAL_IMG_NR="1" ID="PAGE1">
      - <PrintSpace WIDTH="138" HEIGHT="213" VPOS="32" HPOS="17">
        - <TextBlock WIDTH="32" HEIGHT="5" ID="P1_TB2" VPOS="32" HPOS="17">
          - <TextLine WIDTH="30" HEIGHT="3" VPOS="33" HPOS="18">
                                                                    ="1886"/>
                                                                    ="."/>
                                                                    ="13"/>
                                                                    ="."/>
                                                                    T="Januar"/>
                                                                    ="."/>
                                                                    21">
                                                                    T="I"/>
                                                                    T="."/>
                                                                    NT="Jahrgang"/>
                                                                    T="."/>
            <SP WIDTH="4" VPOS="32" HPOS="141"/>
            <String WIDTH="4" HEIGHT="4" VPOS="32" HPOS="145" CONTENT="No"/>
            <String WIDTH="1" HEIGHT="1" VPOS="35" HPOS="150" CONTENT="."/>
            <SP WIDTH="2" VPOS="32" HPOS="150"/>
            <String WIDTH="2" HEIGHT="3" VPOS="32" HPOS="152" CONTENT="2"/>
            <String WIDTH="1" HEIGHT="0" VPOS="35" HPOS="155" CONTENT="."/>
          </TextLine>
        </TextBlock>
```

Lösungsoption I: Kernproblem ist die Erkennung von Frakturschriften



Lösungsoption II: Mustererkennung



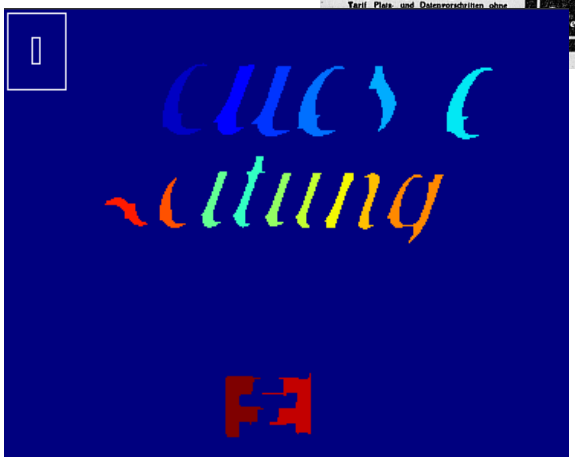
Lösungsoption II: Musterwiederholungen bei periodischen Werken

Lösungsoption II: Ermittlung eines Fingerprints

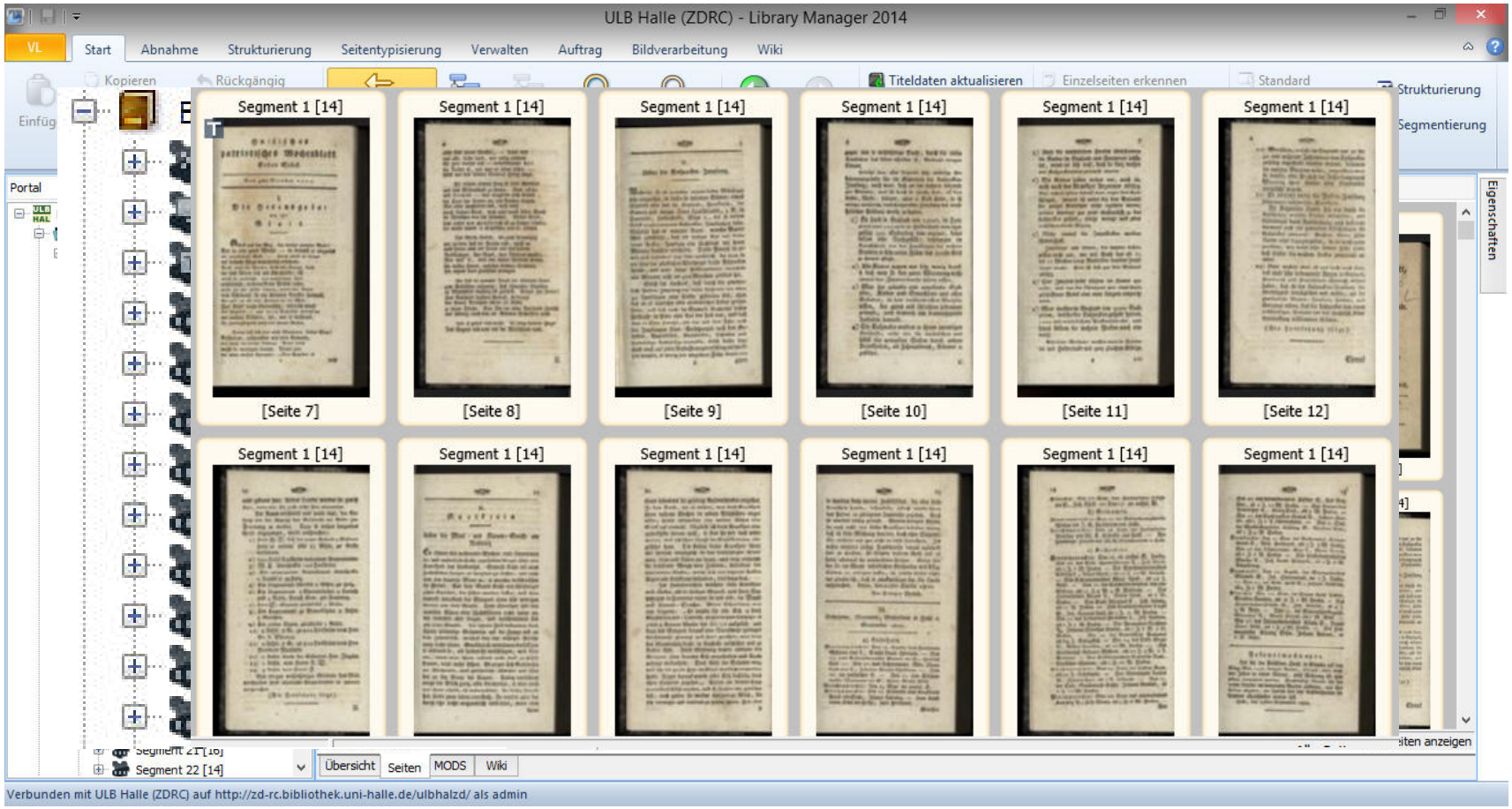
Lösungsoption II: Technische Parameter zur Ermittlung des Fingerprints

- **Ziel**: Identifikation hinreichend eindeutiger Objekte (Pixelhaufen) von bestimmter Größe auf einer Seite, um eine Segmentierung eines Seitenstroms (Jahrgang) in homogene Segmente zur erreichen.

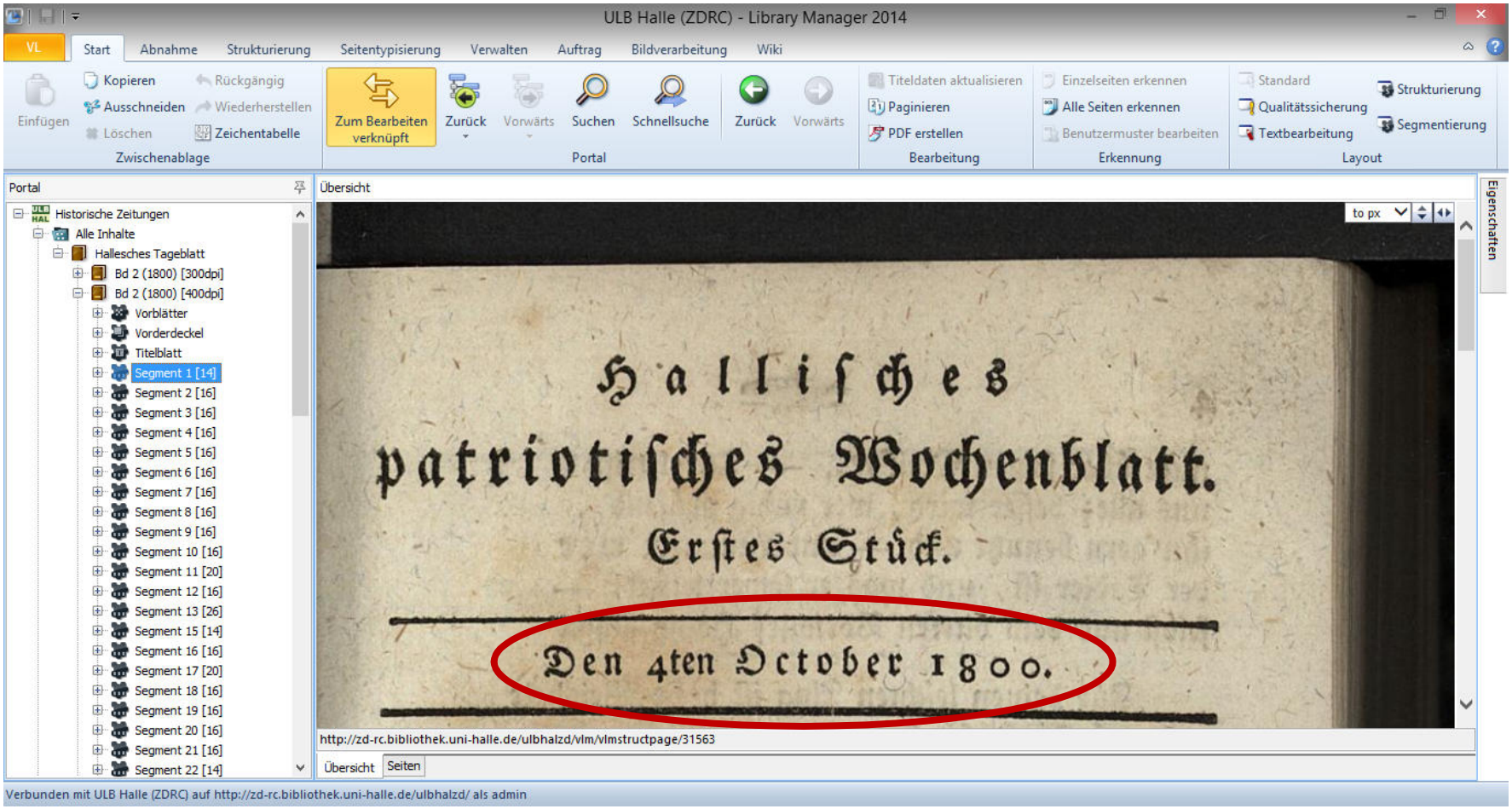
- Objektgröße
- Erosion/Dilation
- Position
- Objektanzahl(min/max)



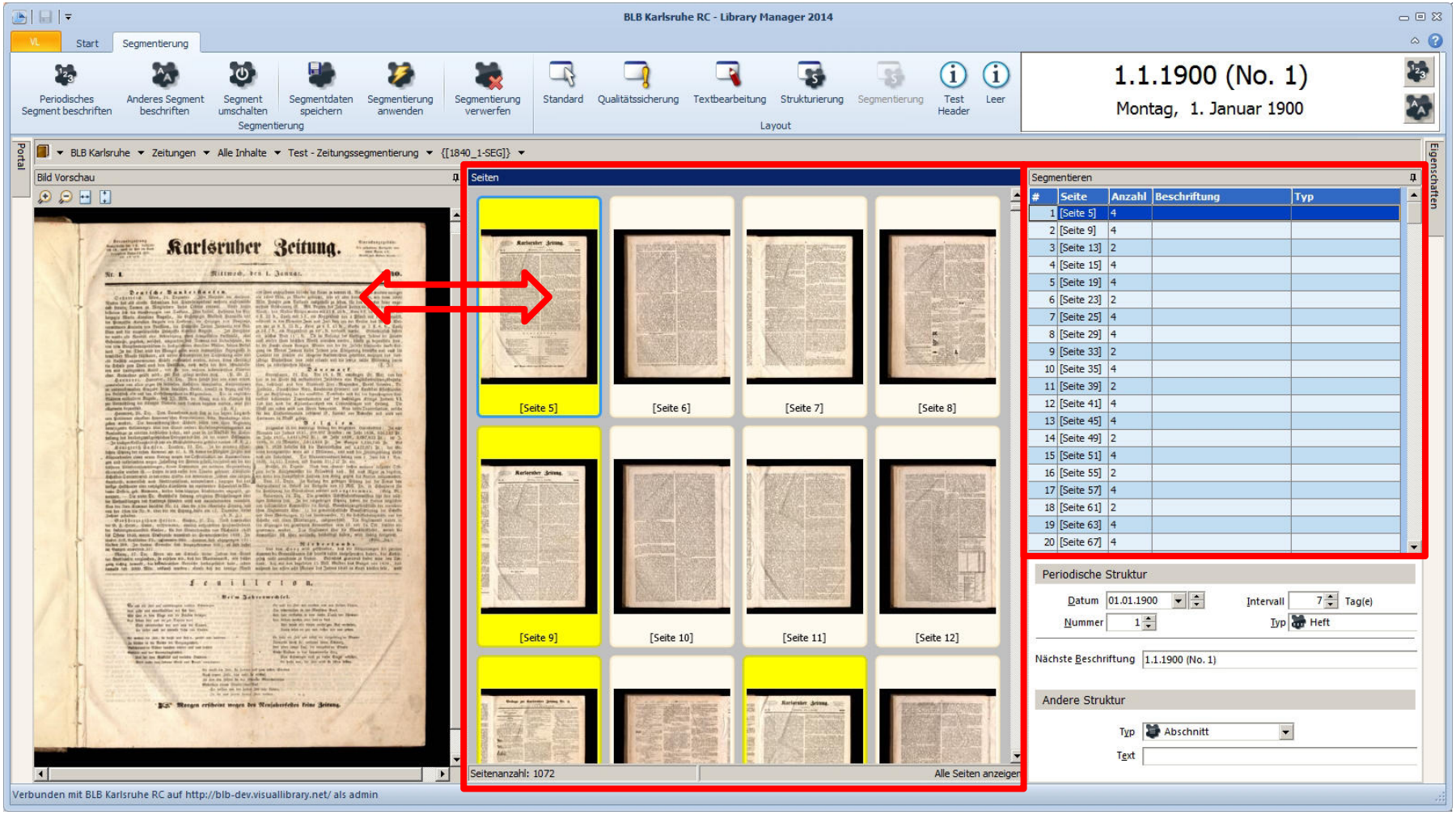
Einsatz in der Projekt-Praxis: Automatische Segmentierung eines Jahrgangs



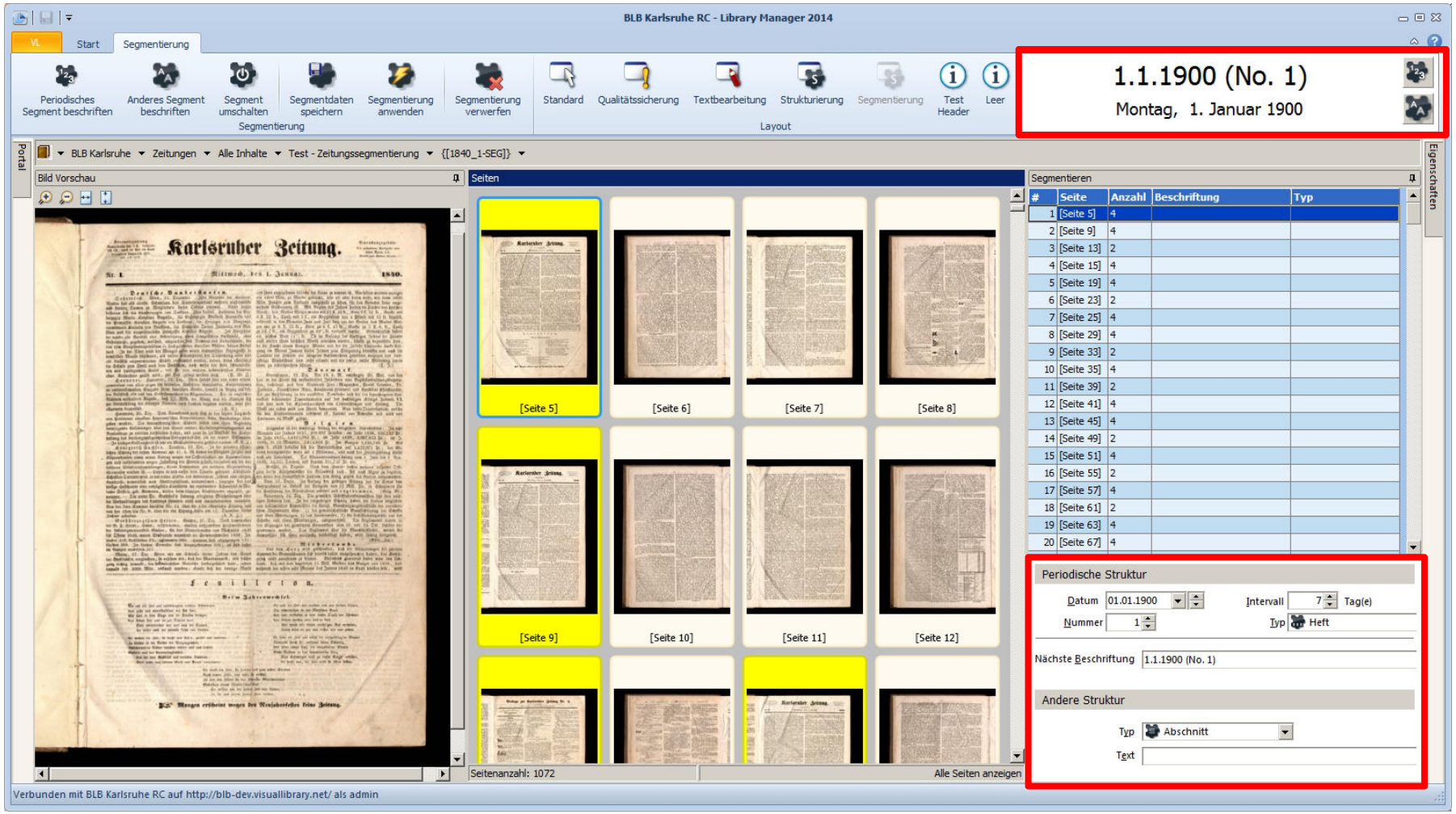
Einsatz in der Projekt-Praxis: Qualitätssicherung der segmentierten Jahrgänge



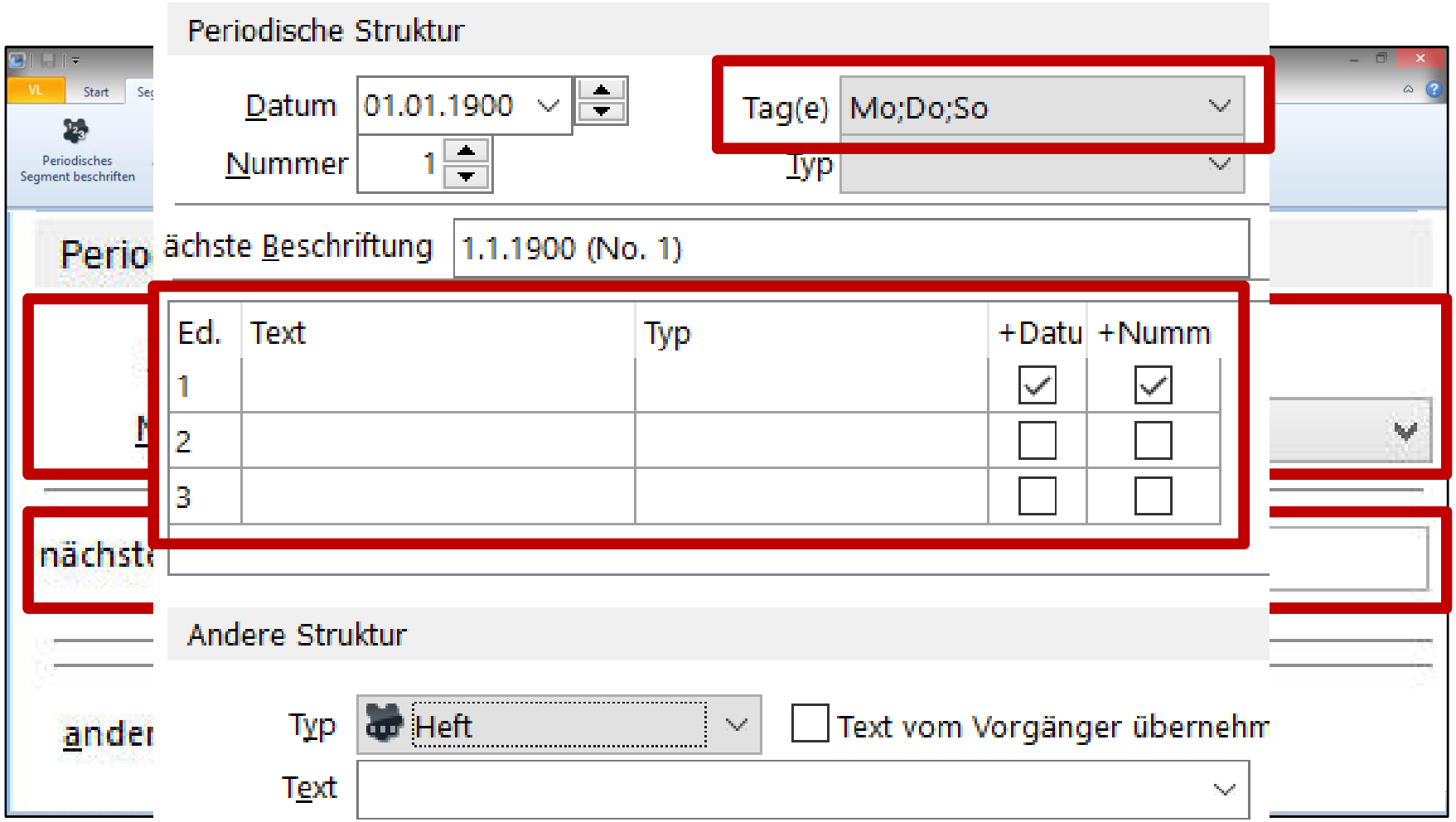
Einsatz in der Projekt-Praxis: Manuelle Bearbeitung



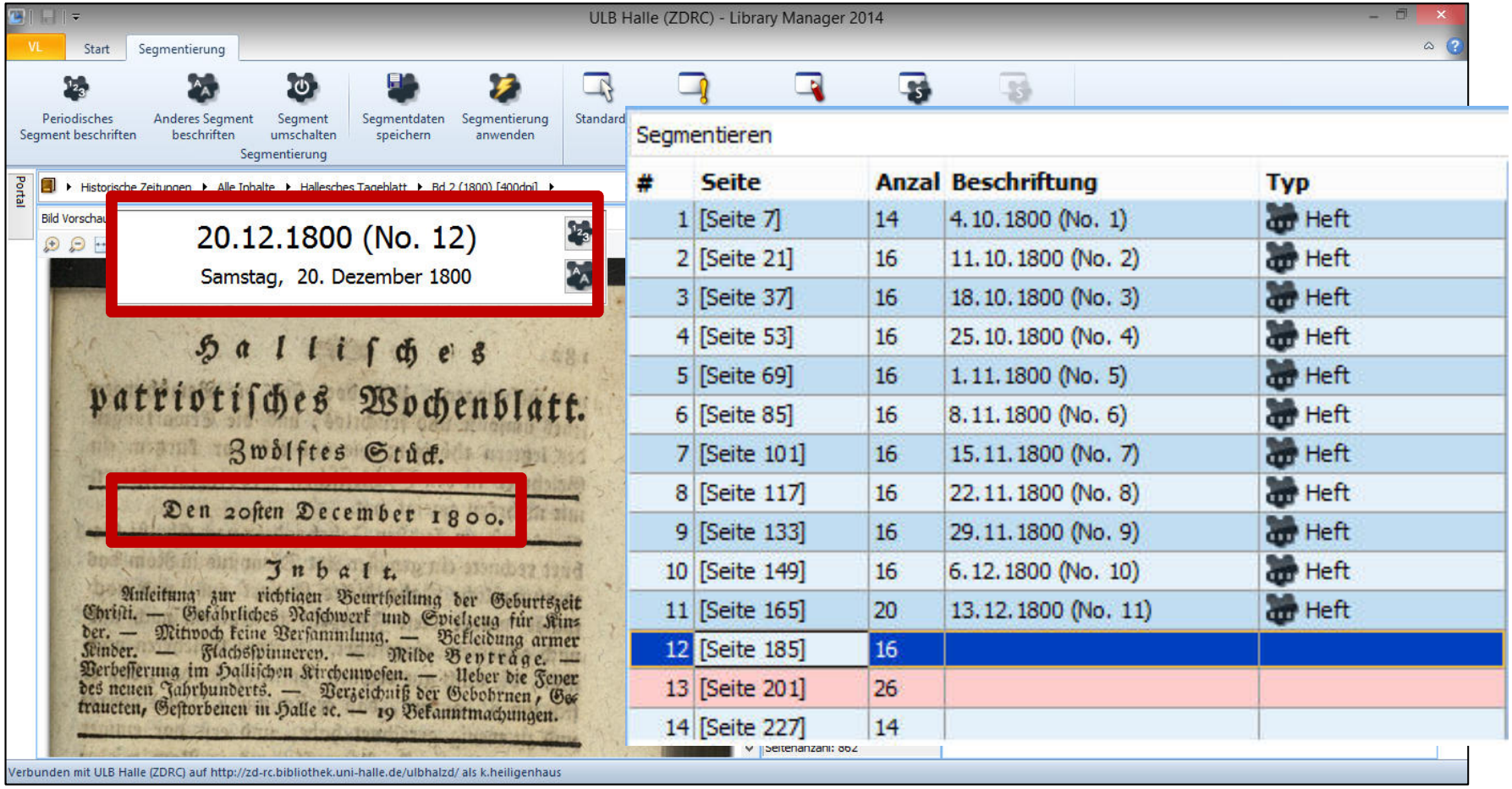
Einsatz in der Projekt-Praxis: Manuelle Bearbeitung



Einsatz in der Projekt-Praxis: Manuelle Bearbeitung



Einsatz in der Projekt-Praxis: Manuelle Bearbeitung



universitätbonn

Rheinische Friedrich-Wilhelms-Universität Bonn

Universitäts- und Landesbibliothek Bonn

Zeitungen | Tag

Sie sind hier: Zeitungen → Jahre → 1789 → 2.4.1789

Zeitungen → Jahre → 1789 → 2.4.1

- Handschriften und Autographen (3394)
- Bibliothek Goussen (928)
- Translation. Die kognitive Aneignung europäischer Schlüsselkonzepte in asiatischen und nahöstlichen Gesellschaften (1860-1945) (22)
- Prinz-Georg-Bibliothek (37)
- Rheinische Landeskunde (807)
- Zeitungen (167)
  - Bonn (32)
  - Rhein-Sieg-Kreis (11)
  - Bergisches Land (24)
  - Kreis Aachen / Düren / Heinsberg (29)
  - Kreis Ahrweiler / Euskirchen / Eifel (19)
  - Kreis Kleve (6)
  - Kreis Neuwied / Westerwald / Mittelrhein (22)
  - Mittelhessen (1)
  - Bad Kreuznach / Rhein-Hunsrück-Kreis / Trier (13)
  - Regierungsbezirk Düsseldorf (7)
  - Regierungsbezirk Köln (1)
  - Saarland (2)

Gazette de Bonn : 2. April 1789

Bonn : [Pauli], 1789

Gnädigst privilegirtes Bönnisches Inte

Staat Köln

Bonn : Intelligenz-Comtoir, 1789

(Nro. LIII.)

GAZETTE DE BONN

AVEC PRIVILEGE DE SON ALTESSE SEREN: ELECT: DE COLOGNE.

DU JEUDI 2. AVRIL 1789.

De Stockholm, *le 10. Mars.*

On a amené ici, hier, fous une forte efcorte de la bourgeoiſie, le colonel Mongommery, les barons de Klingsporre, Hastefko & d'Otter, les Lieutenants-colonels Comtes de Lionſtedt & Enchielm, & le capitaine de Tærre; tous gentils-hommes arrêtés en Finlande. Le Général-major Armfeld eſt tombé malade en chemin, & fera amené ici, dès que ſon état permettra de le tranſporter, avec pluſieurs autres officiers qu'on ne nomme point, & qui ont encore été plus nouvellement arrêtés à Swesbourg.

Il s'en faut bien d'ailleurs que la province de Finlande ait approuvé la conduite de ces officiers; elle a au contraire donné & donne encore plus que jamais des preuves de ſon adhéſion aux volontés du Roi. Les Députés des payſans de cette province ont addreſſé à S. M. un écrit, qu'elle a bien voulu recevoir elle même, par lequel ils l'aſſurent de leur fidélité à toute épreuve. Le Roi a repondu à cette adreſſe avec l'éloquence onctueuſe & pleine de ſentiments, qui caractériſe tout ce qui ſort de ſa bouche.

Les habitans du diſtrict de Hollola ſe diſtinguent encore plus par leur attachement pour le Roi. Ils offrent de garder eux mêmes leur frontières, & de lever pour cet effet 20 ou 30 mille hommes, s'il eſt néceſſaire.

Le Senateur Comte Erie de Stockenſtröm a demandé & obtenû de ſe retirer du Senat. Le Doyen des Senateurs, le Baron de Harmelin eſt mort à Upſale, à l'âge de 82 ans. Il étoit fils du ſécretaire d'état du même nom, qui travailloit ſous les ordres du prémier miniſtre, Comte de Piper, du tems de Charles XII. Il fut fait priſonnier avec lui après la bataille de Pultawa, & mourut avant d'être échangé.

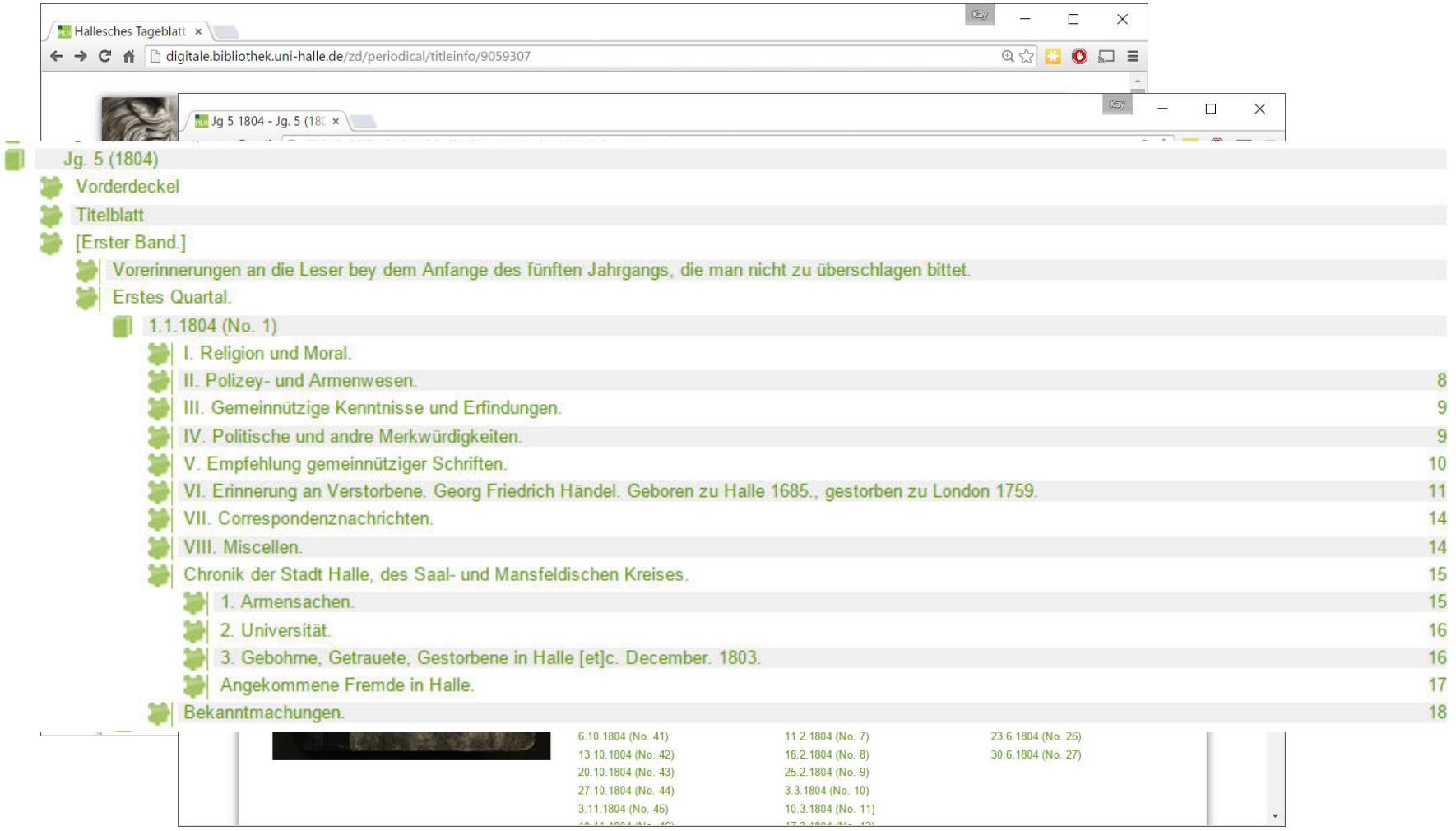
Outre la ſomme dont le Roi a la diſpoſition ſur la banque, elle lui fournira encore pendant tout le tems de la campagne prochaine, ſi elle a lieu, comme il n'y a que trop ſujet de le préſumer, 500 mille écus par mois.

Deux heures avant l'arrivée des officiers arrêtés en Finlande, on avoit reçu en cette ville du même endroit 70 priſonniers ruſſes, parmi leſquels il y avoit pluſieurs prêtres grecs, quelques tartares & différens officiers. Il ſe trouvoit parmi ces priſonniers un eſpion. Le concours du peuple étoit immenſe au moment ou ces deux tranſports ſont entrés dans la ville, & on a ſouhaité à grands cris la bien-venue aux officiers. Pour leur faire place dans le Friéderichs-hoff on avoit d'avance tranſporté les priſonniers nobles

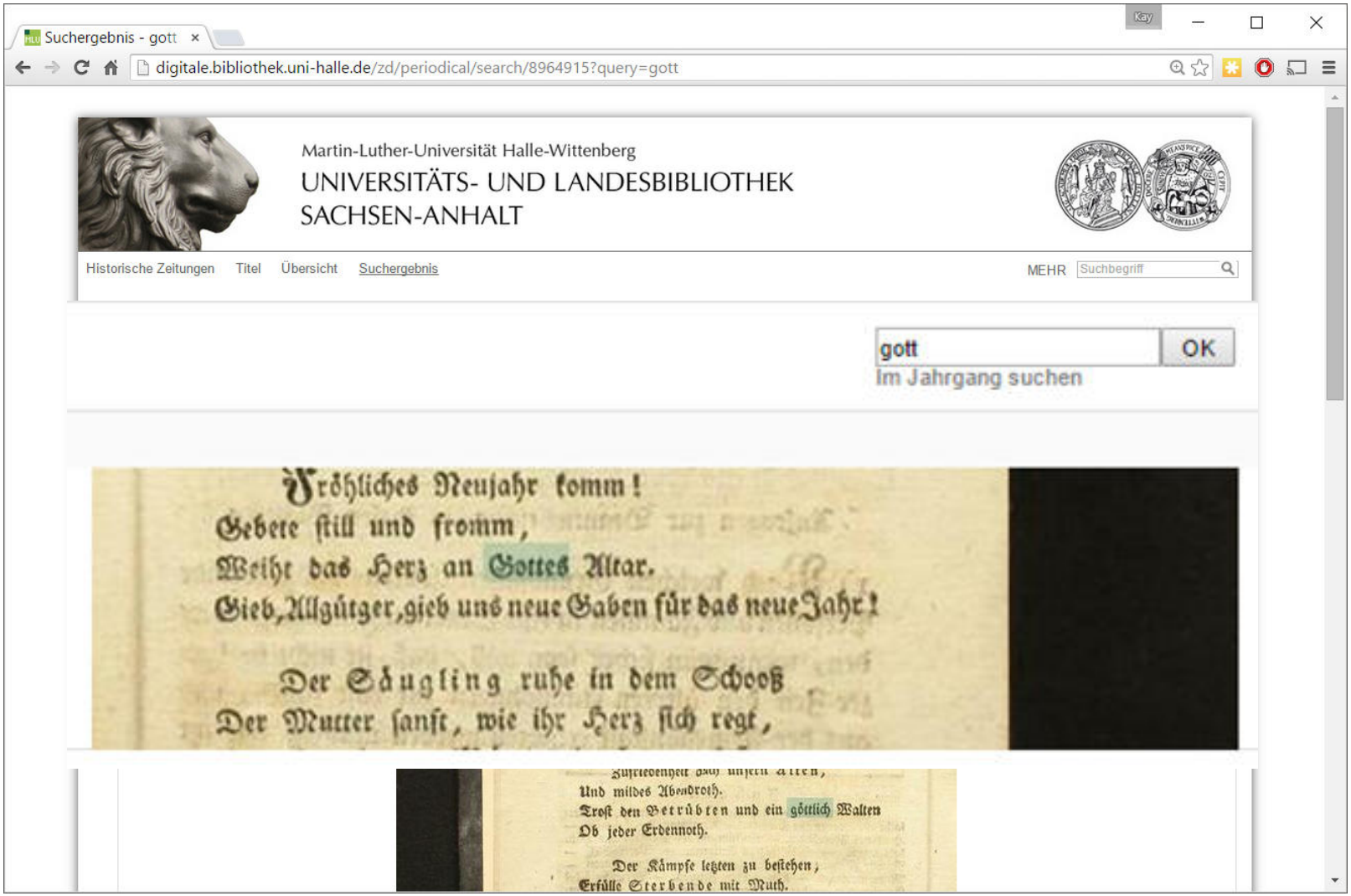
IV.

Vertiefte Erschließung

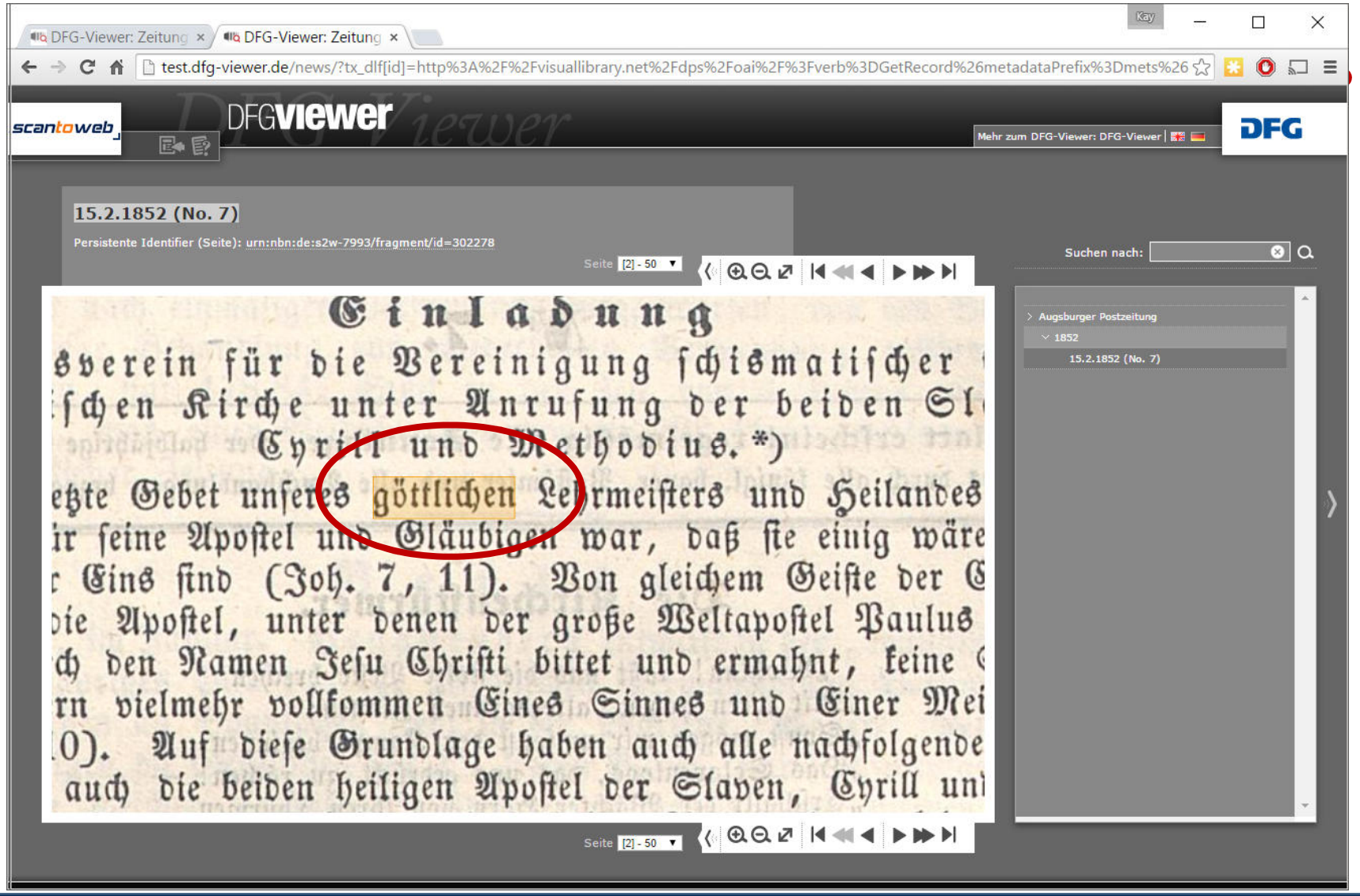
Vertiefte Erschließung: Rubriken und Artikel



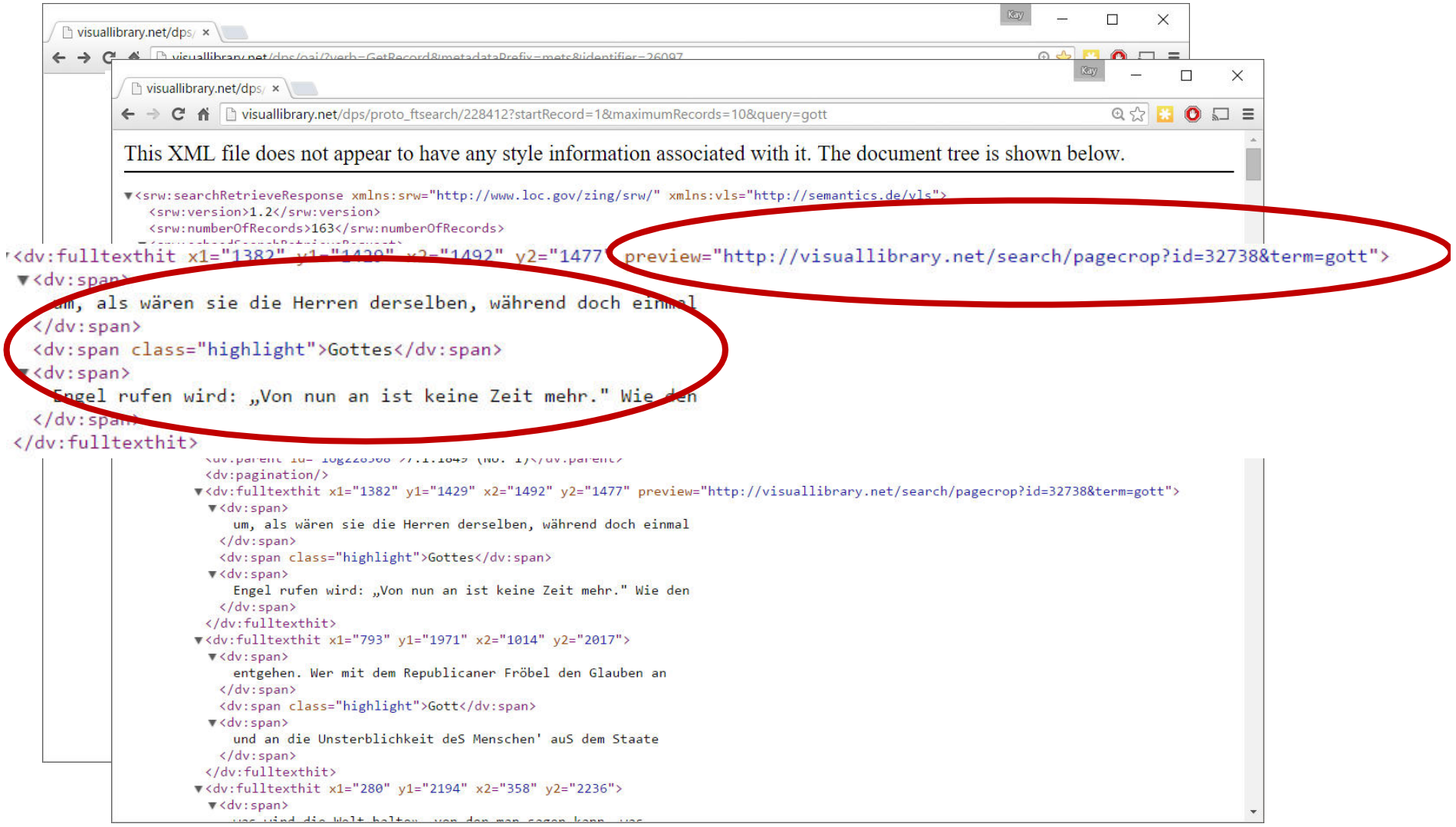
Vertiefte Erschließung: Volltextsuche



Vertiefte Erschließung: Volltextsuche



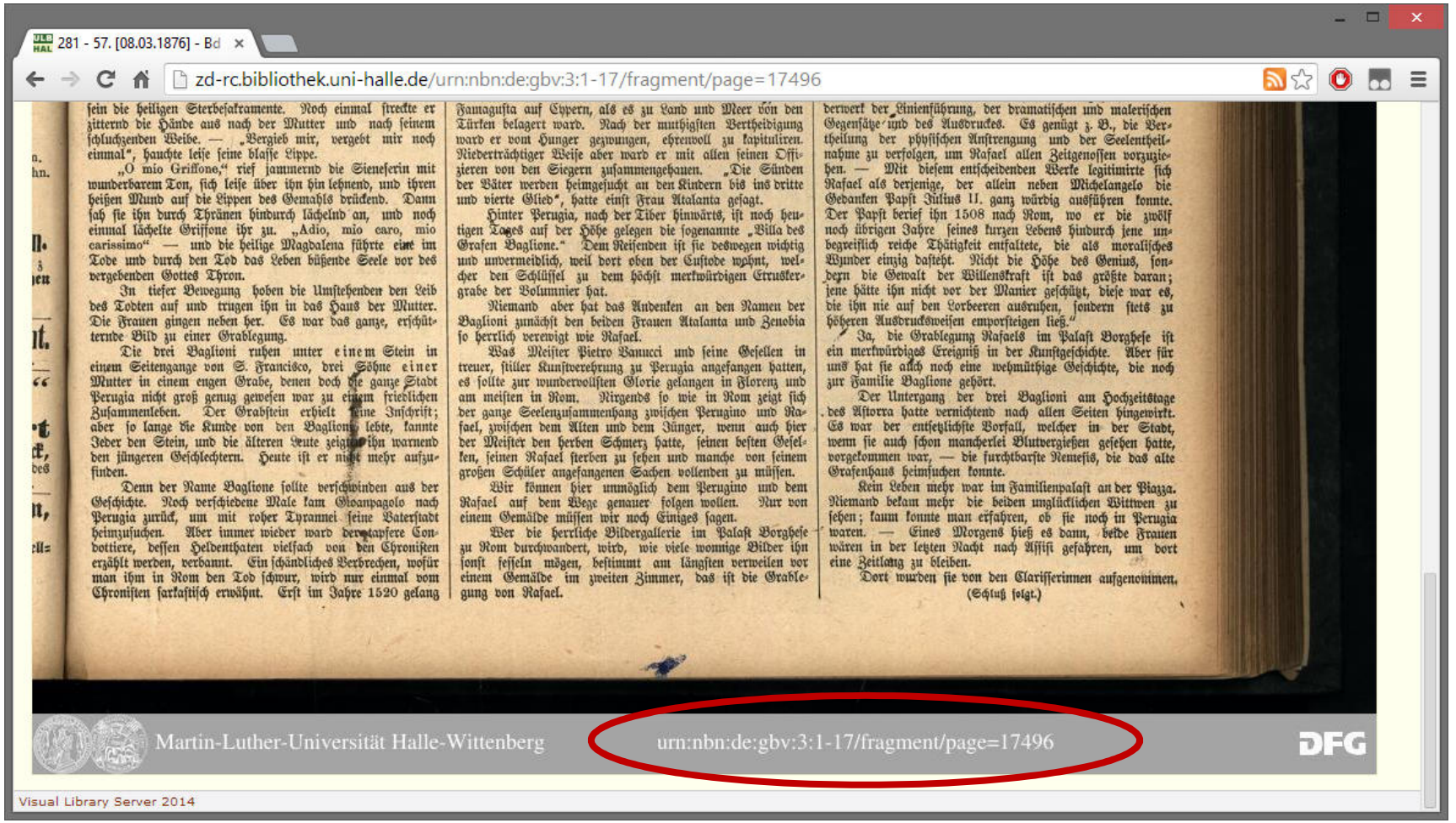
Vertiefte Erschließung: Volltextsuche



V.

Ausblick: Persistente Adressierung

Ausblick: Persistente Adressierung von Einzelseiten



Ausblick: Persistente Adressierung von Artikeln

VI.

Weitere Anwendungs-szenarien

Theaterzettel (am Beispiel der ULB Düsseldorf)

VII.

Zusammenfassung

Zusammenfassung

- Erschließung und Digitalisierung von Zeitungen rücken ihrem Stellenwert gemäß in den Fokus der Öffentlichkeit, Forschung – und Förderung
- Mindestanforderungen der Erschließung und Präsentation erfordern angesichts der Besonderheiten von Zeitungen automatisierte Verfahren
- Mustererkennung, Segmentierungsfunktion und Datenindexierung auf Ausgabenebene von *Visual Library* reduzieren Personal- und Zeitaufwände
- Kalenderbasierte Recherche- und Navigationsoption gewährleisten einen ganzheitlichen, gezielten Zugriff auf gesuchte Einzelausgaben
- Die Volltexterschließung bietet darüberhinaus maximale Recherchefunktionalität für den Nutzer
- Das Verfahren kann – mit entsprechenden Modifikationen – auch für vergleichbare, periodisch erscheinende Werke genutzt werden