

PPS Projekt RHEINPFALZ Verlag und Druckerei GmbH & Co. KG

## Das Altarchiv von 1945 bis 1974 ist digitalisiert und steht demnächst als Web:digiPaper zur Verfügung

PPS PrePress Systeme GmbH erhielt im Dezember 2012 von der **RHEINPFALZ Verlag und Druckerei GmbH & Co. KG** den Auftrag über die Digitalisierung des Zeitungsarchivs von 1945 bis 1974.

### 1. Anlieferung der Zeitungen

Gekennzeichnet war dieser Auftrag von besonderen Schwierigkeiten. Die Zeitungen wurden lose und auf halbes Format gefaltet angeliefert. Zwischen 40 und 60 Jahren lagen Sie in diesen Zustand. Bei diesen Vorlagen war eine Vorverarbeitung nötig, die wir nachfolgend beschreiben und in der Hauptsache bildlich darstellen.



### 2. Die Zeitungen müssen aufgefalt werden.

Durch die jahrzehntelange Lagerung ist im Mittelfalz eine derart große Papierspannung entstanden, die eine Weiterverarbeitung unmöglich machte. Die Spannung musste mit Dampf eliminiert werden, d. h., jede Ausgabe wurde mit einem Dampfbügeleisen glattgebügelt.





Die Zeitungen wurden anschließend nach Ausgabedatum und Ausgabennummer sortiert und mit einer Klebebindung versehen.



### 3. Scannen nach der Klebebindung

Die Bücher wurden je nach Umfang quartalsweise oder monatsweise gebunden. Eine entsprechende Beschriftung der Bände erfolgte, nachdem die Buchblöcke in vorgefertigte Decken eingehängt wurden.

### 4. Aufbereitung der Seiten

Die gescannten Seiten durchlaufen eine intelligente Bildbearbeitung (IBB), die die Schmutzpartikel innerhalb der Seite entfernt und einen sauberen definierten Rand zieht. Anschließend erfolgt die OCR mit einer optimierten Layouterkennung, die für gute Ergebnisse bei der Separierung der Einzelartikel sorgt. Um die Inhalte vollständig weiterzuleiten, speichern wir alle Seiten-Informationen als native XMLs.

## 5. Aufbereitung der Inhalte

Die Automatische-Artikel-Separierung (AAS) – eine von PPS entwickelte Software – strukturiert die aus den FineReader nativ XMLs erzeugten Blöcke über eine logische geometrische/semantische Analyse und ordnet diese den korrekten Lesefluss zu. Ebenso wird Werbung herausgefiltert und Todesanzeigen erkannt und entsprechend getagged.

Desweiteren erkennt die AAS durch eine typografische Analyse Dachzeilen, Haupttitel, Untertitel, Vorspann und den Artikeltext sowie Bildunterschriften und tagged diese entsprechend. Dies wird für die anschließende Aufbereitung für das Web:digiPaper Framework benötigt.

## 6. Aufbereitung msh Web:digiPaper

Aus den AAS XMLs werden die Inhalte von PPS msh Web:digiPaper-konform aufbereitet:

- Skalierung der Vorschau- sowie Detailbilder
- Ausschneiden der Artikelfaksimile sowie Bilder
- Generierung Web:digiPaper konformer Artikel-XMLs
- Berechnung der Koordinaten der Artikelfaksimile und der Bilder zur korrekten Darstellung im Web:digiPaper Framework
- Verlinken aller Elemente in der digiPaper.xml
- Erstellung Web:digiPaper-konformer ZIP-Files zur Übertragung auf das Produktivsystem bei msh