

Drucklegung – MinRProt

Beschreibung des Prozesses der Drucklegung eines Bandes der „Protokolle des Bayerischen Ministerrats 1952-1962“

Inhalt

- [Allgemein](#)
- [Gesamtdatei](#)
- [Registererstellung](#)
- [Vollfinale Fassung](#)
- [Weitere Werkzeuge](#)
- [Grapische Darstellung](#)

Allgemein

Im Folgenden wird der Prozess der Drucklegung des Bands „Kabinett Ehard III/3 1953“ (ehr33) nachvollzogen. Ausgangsbasis sind die XML-Dateien, die zu diesem Band gehören. Das sind zum einen die bandspezifischen Dateien der Protokolle und der Editionsabschnitte. Zum anderen sind dies die Dateien aus der Registerdatenbank (RDB), die bandübergreifend angelegt sind.

Ziel ist die Erstellung eines *Druck-PDFs*, einer PDF-Datei, die den kompletten Bandinhalt in einer druckfertigen Fassung enthält. Das *Druck-PDF* soll direkt an den Verlag geschickt werden können und dort ohne weitere (substantielle) Veränderung für den Druck hergenommen werden können.

Die Umwandlung von XML nach PDF geschieht mittels XSL-FO und dem in Oxygen integrierten Apache FO Prozessors (Apache FOP).

Gesamtdatei

Die *Gesamtdatei* „ehr33_gesamt.xml“ dient zur Zusammenstellung des Druckbandes. In ihr werden alle relevanten XML-Einzeldateien (Editionsabschnitte, Protokolle und Registerdatenbanksdateien) gesammelt und in der Reihenfolge angeordnet, in der sie im fertigen Band stehen sollen. Die *Gesamtdatei* enthält dabei nicht den Inhalt der Dateien selbst, sondern `<xi:include>`-Verweise auf deren `<body>`-Elemente.

Hinter den RDB-Dateien stehen in der Gruppe „Ergänzungen“ noch zwei Erläuterungstexte, die im *Gesamt-PDF* dem Personen- und Orts- und Sachregister vorangestellt werden.

Vorumbuch

Integration

Während die *Integration* gewöhnlich automatisch bei der Umwandlung von *Gesamtdatei* zu *Gesamt-PDF* geschieht, kann mittels des Transformationsszenarios „Gesamtdatei – Integration – XML“ auch einzeln durchgeführt werden.

Ergebnis der Umwandlung, die auf die *Gesamtdatei* angewendet werden kann, ist eine XML-Datei, die den kompletten Text des Editionsbandes, sowie die referenzierten Ausschnitte aller RDB-Dateien enthält. Diese Datei bietet die Möglichkeit, bequem Such- und XPath-Anfragen auf den gesamten Editionstext anzuwenden. So kann zum Beispiel systematisch nach Fehlern, Inkonsistenzen oder fehlerhaften Verweisen gesucht werden.

Mittels des Transformationsszenarios „Gesamtdatei – Druck – PDF“ kann die *integrierte Gesamtdatei* zu Testzwecken in ein *Gesamt-PDF* umgewandelt werden.

Umbruchkorrektur

Alle Verbesserungen, die im Zuge der Drucklegung notwendig werden, müssen prinzipiell in den einzelnen Editionsdateien durchgeführt werden, nicht in der (*integrierten*) *Gesamtdatei*. Um die Änderungen im Druckbild zu sehen, muss anschließend ein neues *Gesamt-PDF* erzeugt werden. Alternativ kann aus einer einzelnen Datei mit dem Transformationsszenario „Einzeldruckvorschau – PDF“ ein PDF erzeugt werden, das dem Drucklayout entspricht. Es kann hier aber, z.B. wegen überhängender Fußnoten, zu Abweichungen im Seitenumbruch gegenüber der Gesamtdateiumwandlung kommen. Deshalb sollten alle Verbesserungen auch einmal im *Gesamt-PDF* überprüft werden.

Seitenumbruch

Mit der Aktion „Druck“ > „Erzwungenen Seitenumbruch einfügen“ kann der Seitenumbruch innerhalb eines Dokuments beeinflusst werden. (Da jedes Dokument in MinRProt auf einer neuen Seite beginnt, haben solche Änderungen keine Auswirkung auf andere Dokumente). Sie fügt eine XML-Processing-Instruction ein, die bei der PDF-Umwandlung den Seitenumbruch an der entsprechenden Stelle bewirkt. *Erzwungene Seitenumbrüche* können nur zwischen Absätzen oder <div>-Elementen eingefügt werden. Unter Umständen muss mit Umbrüchen an verschiedenen Stellen eines Dokuments experimentiert werden, um das beste Druckbild zu erreichen.

Zeilenumbbruch

Derzeit gibt es in unseren Frameworks keine Möglichkeit an einer bestimmten Stelle einen Zeilenumbbruch zu erzwingen. Die Silbentrennung wird von einem automatischen, musterbasierten Algorithmus übernommen. Die Muster basieren auf TeX-Methoden und sind im FOP-Prozessor kodiert. Außerdem ist es wichtig, dass geschützte Leerzeichen im Text benutzt werden, um Zeilenumbüche an unerwünschten Stellen zu vermeiden. (z.B. bei „S. 123). Auch trennt die automatische Silbentrennung unter Umständen an Schrägstrichen („/“). Wo dies verhindert werden soll, kann mit der Aktion „Zeilenumbbruch verbieten“ ein Textausschnitt markiert und geschützt werden.

Das Einfügen von geschützten Leerzeichen und Trennverboten kann auch mittels regulärer Ausdrücke und die Oxygenfunktion „Ersetzen in Dateien“ erfolgen.

Umbruch

Sobald die Drucksatzbearbeitung abgeschlossen ist, besitzt man den Umbruch, der intern auch als *textfinale* Fassung bezeichnet wird. Ab jetzt sollten keine umbruchrelevanten Änderungen am Editionstext mehr durchgeführt werden.

Registererstellung

Sobald die Edition *textfinal* ist, kann mit der Registererstellung fortgefahren werden.

Die Transformationsszenarien für das *Gesamtdokument* können auf Basis der *Registerdatenbank-Referenzen* automatisch die Seitenzahlen ermitteln, auf denen die in den RDB-Dateien enthaltenen Entitäten vorkommen. Dabei filtern sie alle Einträge heraus, für die keine Referenz im Text vorkommen (Ausnahme: in der geschachtelten RDB für Sachschlagwörter, bleiben Einträge erhalten, die selbst keine Referenz im Text besitzen, bei denen aber ein untergeordneter Eintrag über ein verfügt). Im Zuge der Umwandlung werden die Seitenzahlen für das Personenregister (auf Basis von „minrprotRDB_PER.xml“ und das Orts- und Sachregister (auf Basis von

„minrprotRDB_ORT.xml“ und „minrprotRDB_SSW.xml“, die im Zuge der Umwandlung zusammengeführt werden) erzeugt.

Jede RDB-Referenz im Text und alle eingefügten Biogramme werden mit der jeweiligen Seitenzahl wiedergegeben. Der Apache-FOP kann diese Seitenzahlen aber nicht zusammenfassen. Kommt die gleiche RDB-Referenz mehrmals auf der gleichen Seite vor, ist die entsprechende Seitenzahl auch mehrmals aufgeführt. Ebenso wenig werden aufeinanderfolgende Seitenzahlen zusammengefasst.

Rohregister

Um das endgültige Register für den Druckband herzustellen, muss das automatisch erzeugte Register manuell nachbearbeitet werden. Hierfür wird zuerst das Transformationsszenario „Gesamtdatei – Rohregister – PDF“ auf die *Gesamtdatei* angewandt werden. Dieses erzeugt ein *Gesamt-PDF* mit einem angepassten Registerapparat. Das Personen- und das Orts- und Sachregister werden einspaltig statt wie sonst zweispaltig dargestellt. Außerdem sind in jedem Eintrag nach dem Registerbegriff und nach den Seitenzahlen Sonderzeichen eingefügt.

Diese Änderungen gegenüber der Standardansicht erleichtern die automatisierte Weiterverarbeitung. Der hinterlegte Text der Register, wird aus der PDF-Datei extrahiert und mit Hilfe von regulären Ausdrücken in eine einfache XML-Datei zurückgewandelt. In dieser Datei kann nun mit automatischen und manuellen Mitteln die nötige Nachbearbeitung durchgeführt werden, um überflüssige Seitenzahlangaben zu kürzen oder zusammenzufassen.

Druckregister

Aus dem manuell korrigierten Rohregister werden die Druckregisterdateien „ehr33aPSRG“ (Personenregister) und „ehr33aOSRG“ (Orts- und Sachregister) erstellt. Diese enthalten den Text des Registerapparats, wie er im gedruckten Band erscheinen soll. Zusätzlich werden noch *Vortexte* eingefügt mit Erläuterungen zur Registernutzung. Diese ersetzen die Texte aus der Gruppe „Ergänzungen“.

Die Druckregister-Gesamtdatei „ehr33_gesamt_druckregister.xml“ importiert diese neuen Dateien, anstelle der entsprechenden RDB-Dateien.

Über das Transformationsszenario „Gesamtdatei – Druck – PDF“ kann ein neues *Gesamt-PDF* erzeugt werden, das die Druckregister enthält.

Verbesserungen

In den Druckregister-Dateien können letzte Korrekturen an den Registern durchgeführt werden (Änderungen an den Registerbegriffen müssen in den RDB-Dateien nachvollzogen werden). Außerdem besteht die Möglichkeit den Drucksatz zu optimieren. Die Aktion „Erzwungenen Spaltenumbruch einfügen“ fügt eine XML-Processing-Instruction ein, die bei der PDF-Umwandlung an der betreffenden Stelle einen Spaltenwechsel herbeiführt. *Erzwungene Spaltenumbrüche* können zwischen zwei Einträgen oder innerhalb der Seitenzahlen eines Eintrags verwendet werden, um eine möglichst einheitliche Spaltenhöhe zu gewährleisten.

Vollfinale Fassung

Sobald die Druckregister-Dateien fertig verbessert und gesetzt sind, steht die *vollfinale* Version des Bandes. Das endgültige *Druck-PDF* kann nun erzeugt werden, indem man das Transformationsszenario „Gesamtdatei – Druck – PDF“ auf die *Gesamtdatei mit Druckregister* anwendet.

Weitere Werkzeuge

Diff-PDF

Während des Drucklegungsprozesses sollten die Ergebnisse PDF-Umwandlung immer wieder mit einem PDF-Vergleichstool überprüft werden. Wir verwenden hierfür das Programm „diffpdf“ (siehe <http://vslavik.github.io/diff-pdf/>). Es kann visuelle und inhaltliche Unterschiede zwischen zwei PDF-Dateien feststellen.

Mittels solcher Vergleiche lässt sich sicherstellen, dass Änderungen am Editionstext oder den Umwandlungsskripten die erwarteten Auswirkungen haben und keine unerwünschten Nebeneffekte auftreten.

XSL-Fassung

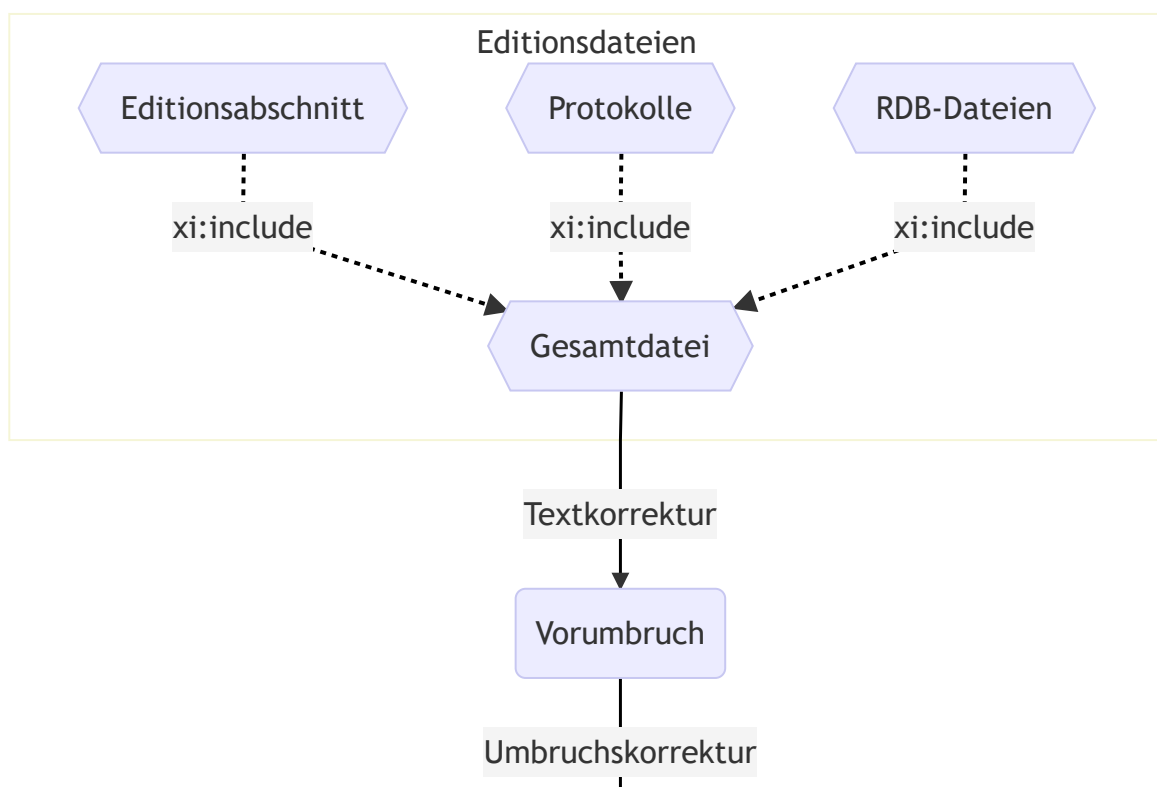
Mit dem Transformationsszenario „Gesamtdatei – Druck – XSL“ kann die XSL-FO-Version des *Gesamt-PDFs* erzeugt werden, bevor dieses in vom FO-Prozessor zum PDF umgewandelt wird. In diesem Format kann Problemen beim der PDF-Transformation nachgegangen werden.

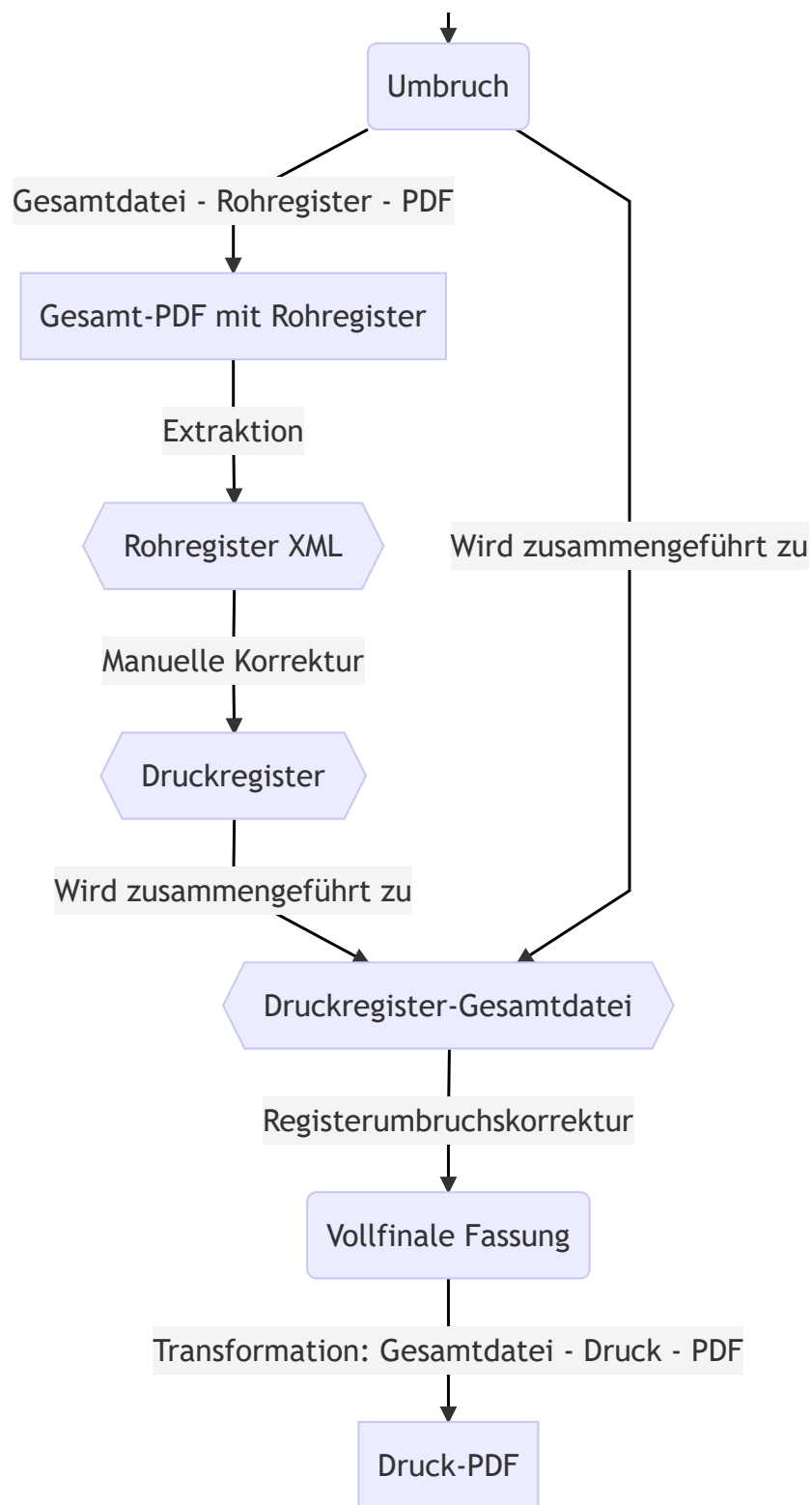
Weitere Verbesserungsmöglichkeiten

Neben den Aktionen zum Einfügen von *erzwungenen Seiten- und Spaltenumbrüchen*, kommen in Spezialsituationen weitere manuelle Eingriffe in die Editionsdokumente vor, um das Druckbild zu bewirken. Diese werden nur in bestimmten Ausnahmefällen eingesetzt und sollen nach Möglichkeit mittelfristig durch systematische Lösungen ersetzt werden. Es gibt für sie keine Aktionen in der Editions Umgebung, sie müssen bei Bedarf manuell im Text-Modus hinzugefügt werden. Ein Übersicht über diese Elemente befindet sich in den [digitalen Editionsrichtlinien zur Edition unter dem Punkt „Druckanpassungen“](#).

Graphische Darstellung

Folgenden Flowchart visualisiert den Ablauf der Drucklegung:





Legende



