



**Staatsbibliothek  
zu Berlin**  
Preußischer Kulturbesitz

Staatsbibliothek zu Berlin

# Zwischen Konzept und Alltag:

## Ein Werkstattbericht zur Neuausrichtung der Zeitungsdigitalisierung an der SBB

Maria Federbusch, Katharina Droese, Lole Westedt und Armin Möller

- Ausgangslage
- Neuausrichtung Zeitungsdigitalisierung – Ziele
- Konzeption Digitalisierungsworkflow für Zeitungen
- Kitodo-Workflow für Zeitungen in der Praxis
- Ausblick



- Digitalisierung von Zeitungen an SBB seit vielen Jahren: ZEFYS-Daten: 678.000 Ausgaben, 6,5 Mio Images
- Eigenentwicklung bzgl. Workflow, Formaten und Präsentation – fehlende Schnittstellen
- Bisheriger ZEFYS-WF: Digitalisierung mit Sara-Tool; aufwendige Datenprozessierungen, manuelle Paketablieferungen ausgewählter Zeitungen an DDB
- Notwendigkeit der Umstellung (inklusive Workflow, der alle entstehenden Daten verwaltet und Formatvorgaben der DDB folgt)



Startseite / Liste der digitalisierten Zeitungen DE / EN

Alle Zeitungstitel

A B C D E F G H I K L M N O P R S T U V W Z

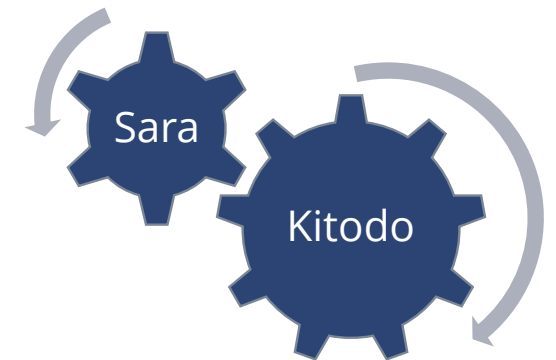
Zeitungstitel filtern

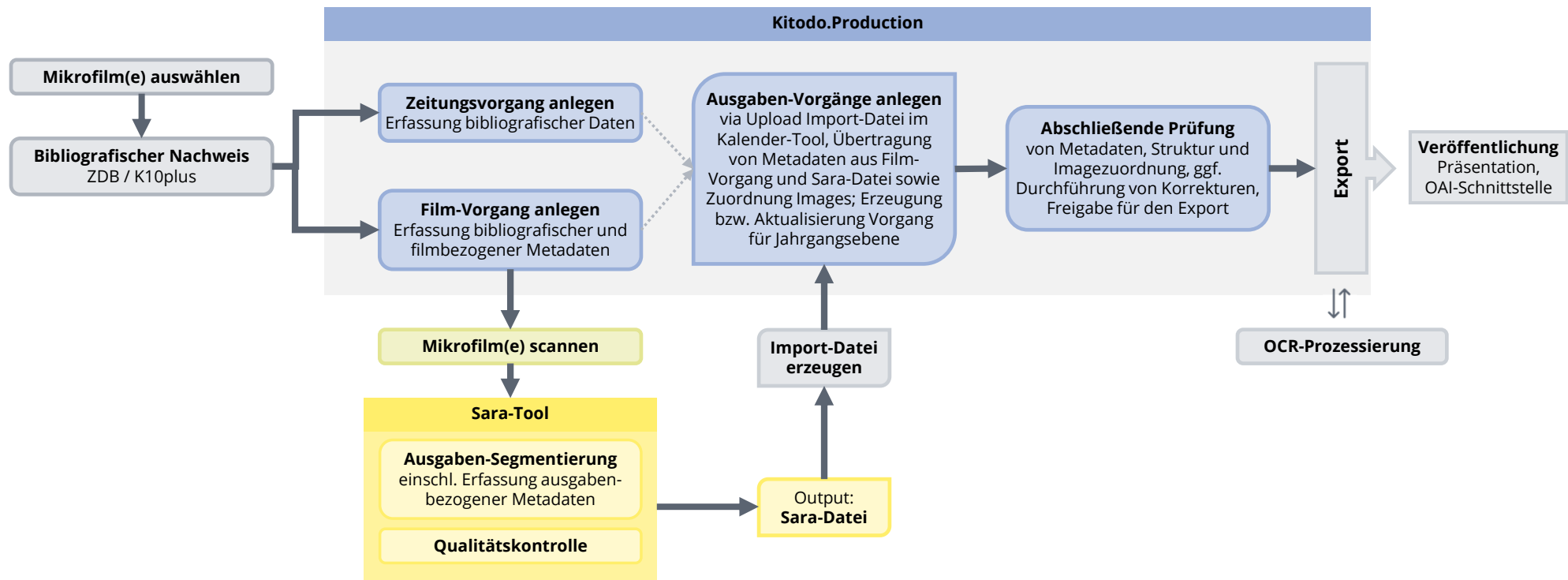
- A
  - Ägyptische Nachrichten 1912
  - Allgemeine Zeitung 1857 - 1858
  - Ämtliches Kreisblatt für den Kreis Koschmin 1909 - 1918
  - Ämtliches Kreisblatt und Anzeiger für den Kreis und die Stadt Koschmin 1905 - 1909
  - Der Anzeiger 1799 - 1802
  - Anzeiger für das Havelland 1905 - 1908
  - Argentinisches Wochenblatt 1942 - 1946
  - Azərbaycan 1943 - 1944
- B
  - Baruther Anzeiger 1925 - 1942
  - Baruther Heimatland 1932 - 1939
  - Der Beobachter an der Weichsel 1806



- Anknüpfung an etablierte Digitalisierungsworkflows bei gleichzeitiger Berücksichtigung materialspezifischer Anforderungen (überw. Rollfilm-Digitalisierung, Granularität auf Ausgabenebene)
- Ausschöpfung von Automatisierungspotenzialen zur Effizienzsteigerung, v.a. im Hinblick auf Massendigitalisierung
- Bereitstellung standardisierter, interoperabler Daten für übergreifende Portale (DDB, Europeana etc.)
- Möglichkeiten für gezielte Korrekturen und Neuexporte schaffen

- Kitodo.Production für Workflow-Steuerung
  - Goobi / Kitodo an der Stabi im Einsatz seit 2008
  - Aufbau und Konfiguration einer Kitodo-Instanz für Zeitungen (Version 3.9)
- Verknüpfung des Einsatzes von Segmentierungswerkzeug (Sara) und Kitodo-Funktionalitäten für Zeitungen (Kalender-Tableau)
- Metadatenerfassung und -zuordnung auf passender Ebene
  - Zeitung: bibliografische Daten (Titel, ZDB-ID, Erscheinungsort, Verlag etc.)
  - Rollfilm: Signatur, Besitzer u.ä.
  - Zeitungsausgabe: Datum, Ausgabennummer, Hauptblatt/Beilage etc.
- automatisierte bzw. skriptbasierte Zusammenführung und Prozessierung der Daten
- gleiches Workflow-Prinzip für Digitalisierung von Mikrofilm und Original





Es beginnt mit einer Filmrolle...

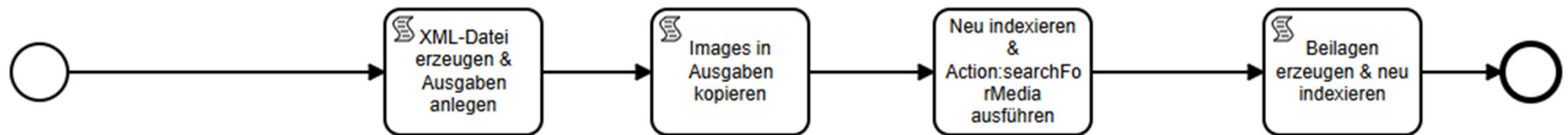


```
10:19:35_root@lx0311:/mnt/zefys/scan_data/POST$ ll
insgesamt 208
drwxrwxr-x 5 b-rr103 529    80 23. Okt 21:12 00001
-rw-rw-rw- 1 b-rr103 529 20085 28. Okt 09:01 00001.cfg
-rw-r--r-- 1 root    529 62207 23. Okt 21:12 00001.filelist.txt
drwxrwxr-x 5 b-rr103 529    80  8. Dez 14:00 00002
-rw-rw-rw- 1 b-rr103 529 16788 15. Dez 09:31 00002.cfg
-rw-r--r-- 1 root    529 47411  8. Dez 14:00 00002.filelist.txt
-rw-rw-rw- 1 b-rr103 529   206  5. Dez 09:50 _Ausgaben_ fehlen_ oder_ beschaedigt_ b-rr103.txt
-rw-rw-rw- 1 b-rr103 529   191  5. Dez 09:50 Ausgaben_ fehlen_ oder_ beschaedigt_ b-rr103.txt
```

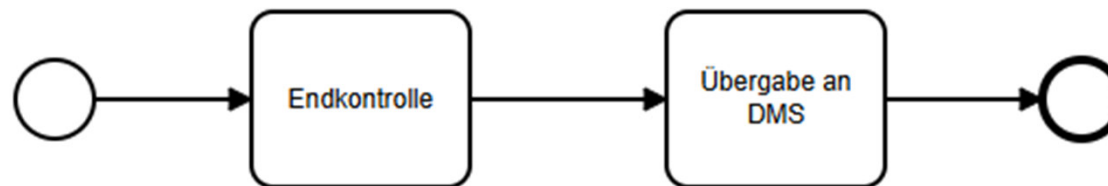
### Sara-Datei: Aufteilung der Filmrolle in einzelne Ausgaben (inkl. Beilagen)

```
10:19:36_root@lx0311:/mnt/zefys/scan_data/POST$ more 00001.cfg
version:2;struct:first_frame,first_thumbnail,date,volume,issue_label
1=gray00001.jpg;gray00001_t.jpg;1866-08-01;1;1;;H;1;0;1;0;0;
5=gray00005.jpg;gray00005_t.jpg;1866-08-02;1;2;;H;1;0;1;0;0;
9=gray00009.jpg;gray00009_t.jpg;1866-08-03;1;3;;H;1;0;1;0;0;
13=gray00013.jpg;gray00013_t.jpg;1866-08-04;1;4;;H;1;0;1;0;0;
17=gray00017.jpg;gray00017_t.jpg;1866-08-04;1;4;;B;1;0;5;0;0;
19=gray00019.jpg;gray00019_t.jpg;1866-08-05;1;5;;H;1;0;1;0;0;
23=gray00023.jpg;gray00023_t.jpg;1866-08-06;1;6;;H;1;0;1;0;0;
```

## Kitodo: Workflow der Filmrolle



## Kitodo: Workflow der Ausgaben



### Skriptaufrufe

Erzeugen der XML-Datei zum Anlegen der Ausgaben:

```
/kitodo/scripts/Sara2Kalender.pl -m (metaFile) -s (generatorsourcepath) -d  
/mnt/zefys/scan_data/ --kalender
```

Verteilen der Images auf die Vorgänge und Anreichern der Metadaten:

```
/kitodo/scripts/Sara2Kalender.pl -m (metaFile) -z (processtitle) -s  
(generatorsourcepath) -d /mnt/zefys/scan_data/ --images
```

Erzeugen und Zuordnen der Beilagen:

```
/kitodo/scripts/Sara2Kalender.pl -m (metaFile) -s (generatorsourcepath) -d  
/mnt/zefys/scan_data/ --beilagen
```

## Kitodo: Anlegen der Filmrolle

The screenshot shows the Kitodo PRODUCTION interface for creating a new production process (Vorgang) based on the 'SBB\_Filmrolle' template. The interface is split into two main sections: 'Basisdaten' (Basic Data) and 'Metadaten' (Metadata).

**Basisdaten:**

- Vorgangstitel \*: 3218487-6-1930104146\_00002
- Dokumenttyp: Filmrolle
- Regelsatz: Zeitungen\_Filmrollen\_Regelsatz
- Vorgangsvorlage:
- Tif-Header Documentname: 3218487-6-1930104146\_00002
- Tif-Header Image-Beschreibung: |<DOC\_TYPE>|<HAUPTTITEL>Die Post|<AUTOREN/HERAUSGEBER>|<JAHR>|<ERSCHEINUNGSORT>|<VERZ\_?>
- geschätzte Seitenanzahl: 0

**Metadaten:**

- Signatur der Quelle: \* Ztg 719 MR-1866, Aug./Dez.
- Film-Nummer Sara-Datei: \* 00002
- Lückenergänzung?: \* Nein
- Digitale Kollektionen: \*
  - Digitale Kollektion: \*  Zeitung
- Zugangstatus: Open Access
- Zugriffslizenz: \* PDM 1.0
- Förderer:  Deutsche Forschungsgemeinschaft
- Digitale Herkunft: \* \_digitizedMicrofilm
- PPN (digital): 1930104146

Buttons at the top right: Abbrechen, Datenübernahme, Speichern + Neu, Speichern.





Buttons at the bottom right: Impressum, Datenschutz, Nutzungsbedingungen, KITODO.PRODUCTION Version 3.9.0

## Kitodo: Erzeugen der XML-Datei für das Anlegen der Ausgaben

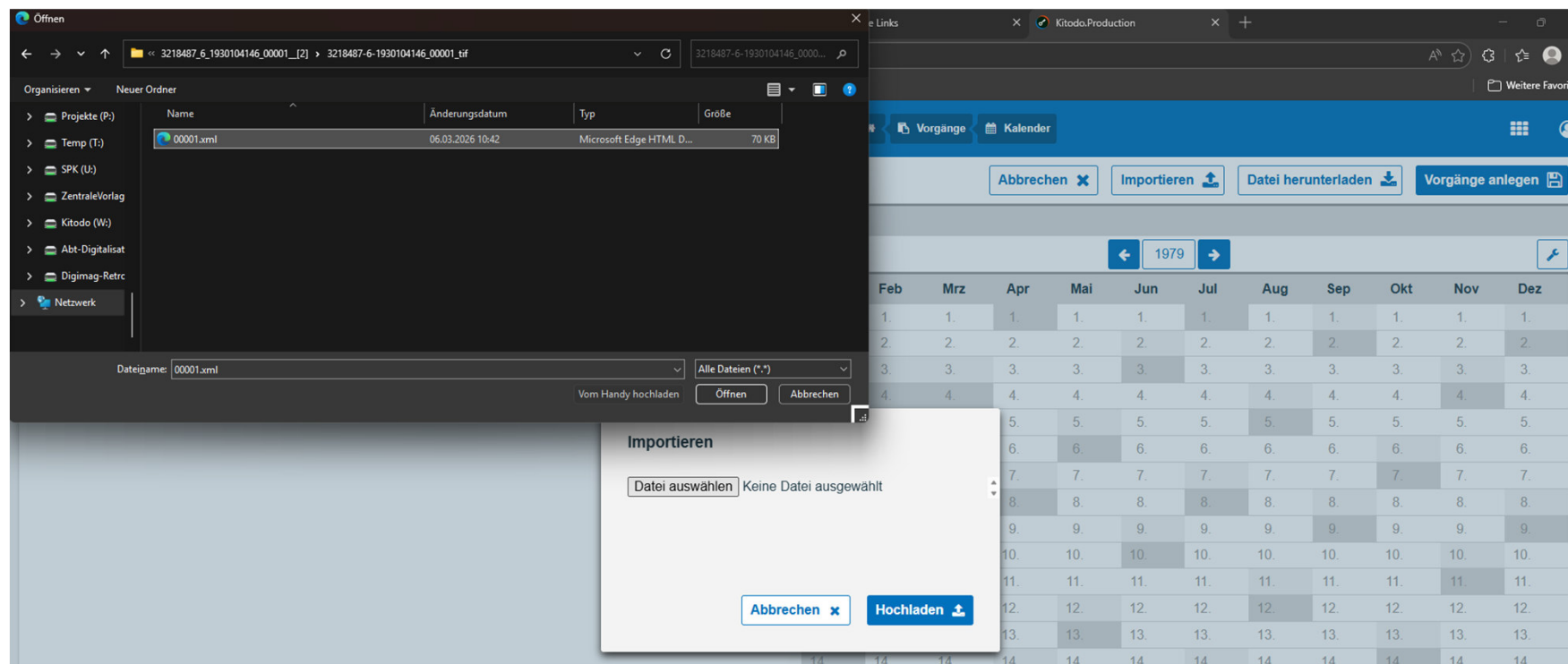
### Aufgabendetails

Aufgabendetails	Kommentare	
Titel	Vorgangstitel	
XML-Datei erzeugen & Ausgaben anlegen	3218487-6-1930104146_00001	

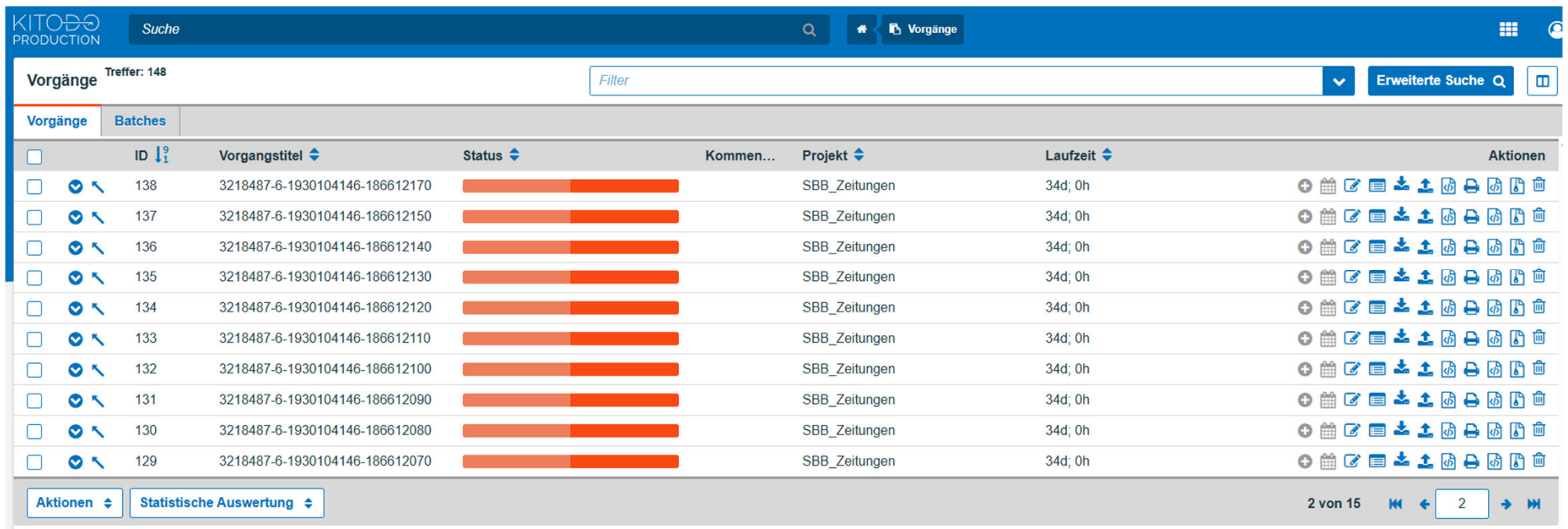
### Aktionen

-  [Skript ausführen: Kalenderdatei erzeugen](#)
-  [Metadaten bearbeiten](#)
-  [Die Bearbeitung dieser Aufgabe wieder abgeben](#)
-  [Die Bearbeitung dieser Aufgabe abschließen](#)

## Kitodo: Ausgaben anlegen mittels XML-Datei-Import



## Kitodo: Ausgaben anlegen mittels XML-Datei-Import



The screenshot shows the Kitodo PRODUCTION interface. At the top, there is a search bar with the text 'Suche' and a search icon. Below the search bar, the main content area is titled 'Vorgänge' with a sub-header 'Treffer: 148'. A 'Filter' input field is present. The main table has two tabs: 'Vorgänge' (selected) and 'Batches'. The table columns are: 'ID', 'Vorgangstitel', 'Status', 'Kommen...', 'Projekt', 'Laufzeit', and 'Aktionen'. The table contains 10 rows of data, each representing a newspaper issue. The 'Status' column shows a progress bar for each issue. The 'Projekt' column is 'SBB\_Zeitungen' for all. The 'Laufzeit' column is '34d; 0h' for all. The 'Aktionen' column contains a set of icons for each row. At the bottom of the table, there are buttons for 'Aktionen' and 'Statistische Auswertung'. The bottom right corner shows '2 von 15' and navigation arrows.

ID	Vorgangstitel	Status	Kommen...	Projekt	Laufzeit	Aktionen
138	3218487-6-1930104146-186612170	Progress bar		SBB_Zeitungen	34d; 0h	Icons
137	3218487-6-1930104146-186612150	Progress bar		SBB_Zeitungen	34d; 0h	Icons
136	3218487-6-1930104146-186612140	Progress bar		SBB_Zeitungen	34d; 0h	Icons
135	3218487-6-1930104146-186612130	Progress bar		SBB_Zeitungen	34d; 0h	Icons
134	3218487-6-1930104146-186612120	Progress bar		SBB_Zeitungen	34d; 0h	Icons
133	3218487-6-1930104146-186612110	Progress bar		SBB_Zeitungen	34d; 0h	Icons
132	3218487-6-1930104146-186612100	Progress bar		SBB_Zeitungen	34d; 0h	Icons
131	3218487-6-1930104146-186612090	Progress bar		SBB_Zeitungen	34d; 0h	Icons
130	3218487-6-1930104146-186612080	Progress bar		SBB_Zeitungen	34d; 0h	Icons
129	3218487-6-1930104146-186612070	Progress bar		SBB_Zeitungen	34d; 0h	Icons

- Aufgabe abgeben, danach automatischer Skriptstart: Images in Ausgaben kopieren
- nächste Aufgabe annehmen: Neu indexieren & action:searchForMedia für alle angelegten Ausgaben ausführen
- Aufgabe abgeben, danach automatischer Skriptstart: Beilagen anlegen
- Als Abschluss neu indexieren
- Ggf. Skript generatImages ausführen

# Kitodo-Workflow für Zeitungen in der Praxis

The screenshot displays the Kitodo-PRODUCTION interface for a newspaper record. The top navigation bar includes 'Suche', 'Vorgänge', and 'Metadaten bearbeiten'. The main header shows the record ID '3218487-6-1930104146-186612290 (ID: 149) (8 Medien)' and action buttons like 'Schließen', 'Validieren', 'Medien umbenennen', 'Speichern', and 'Speichern & schließen'.

The interface is divided into three main sections:

- Strukturdaten:** A tree view on the left showing the hierarchy: [1] Zeitung - 3218487-6-1930104, [3] Zeitungsjahrgang - 3218487, [3] Monat (Zeitung) - 3218487, [3] Kalendertag (Zeitung) - 3218487, [3] Heft/Ausgabe (Zeitung) - 3218487. Below this are options for 'Aufteilung ohne Typ' and 'Heft/Ausgabe (Zeitungs-Ausgabe)'. A 'Beilage' section lists items 5 through 4.
- Metadaten:** A central panel with various fields:
  - Zugangstatus:** Includes 'Open Access' (checked).
  - Zugriffslizenz:** Set to 'PDM 1.0'.
  - Förderer:** 'Deutsche For...' (partially visible).
  - Digitale Herkunft:** '\_digitizedM...'
  - PPN (digital):** '1930104146'.
  - ZDB-ID digital vom Host:** '3218487-6'.
  - ZDB-ID vom Host (Vorlage):** '22985-4'.
  - Ausgabenteil:** (empty field).
- Bild 1, Seite -:** A large preview window showing a scanned page of 'Die Post' newspaper. The page is titled 'Separat-Ausgabe. Jahrgang I. Berlin, Sonnabend den 29. December 1906.' The main headline reads 'Die Post.' Below the headline, there are several columns of text, including a section titled 'Telegraphische Depeschen' and another titled 'Die gute Stadt Hamburg und die Gemeindefähigkeit ihres Senats.' The page also features a small circular logo at the top left and a decorative border.

At the bottom right of the interface, there are links for 'Impressum', 'Datenschutz', and 'Nutzungsbedingungen', along with the version information 'KITODO.PRODUCTION Version 3.9.0'.

Letzte Schritte der Ausgabenvorgänge:

- Berechnen aller Derivate für die Ausgaben
- Endkontrolle der Ausgaben durchführen
- Export (Alle Ausgaben in einen Batch zusammenfassen und daraus exportieren)

Woran wir noch arbeiten:

- Content Server (damit alle Dateien über eine URL erreichbar sind, DFG-Viewer)
- Archivkonzept (Filme oder Ausgaben archivieren?)
- Zeitungsfähige Präsentation

- Praxiseinblick: Vorführung des Sara-Tools durch R. Rüdiger auf dem Best Practice Marketplace im Anschluss
- Herausforderungen und ToDos in der SBB-Zeitungsdigitalisierung:
  - Migration der Zefysaltdaten (Fehlerkorrekturen, Formatanpassungen, Dublettenkontrolle, Neuexport der METS-Daten und Erstellung von iiif-Manifesten)
  - OAI-Schnittstelle für Zeitungen und neue Präsentation
  - Datenfluss ins Deutsche Zeitungsportal; Änderungen nachhaltig vornehmen (Erreichbarkeit der Images erhalten)
  - Zukünftige Entwicklungen (Ressourcen vorausgesetzt) mitdenken: Volltexte, Layouterkennung, Ausgabensegmentierung, Artikelsegmentierung – Überführung von experimentellen Ideen in praxistaugliche Workflowschritte
  - Basis gelegt für spätere Weiterentwicklungen und zusätzliche Prozessierungsschritte

Staatsbibliothek  
zu Berlin  
Preußischer Kulturbesitz

MENSCH. MASCHINE. KULTUR

Startseite Aktuelles Ergebnisse Kontakt

Eynollah ●○○○○○○○

Legend:  
■ Background  
■ Text-Block  
■ Image  
■ Page-Header  
■ Page-Footer  
■ Table

Layoutanalyse historischer Texte mit Deep Learning und Heuristiken

Die Layoutanalyse mit Eynollah erkennt bis zu zehn verschiedene Elemente bzw. Inhaltsbereiche in historischen Dokumenten. Um die Qualität der anschließenden Texterkennung (OCR) zu verbessern, identifiziert Eynollah außerdem die korrekte Reihenfolge der Textregionen (Reading Order Detection).

Die Software, sowie weitere Informationen dazu finden Sie [hier](#).



**Staatsbibliothek  
zu Berlin**  
Preußischer Kulturbesitz

**Katharina Droese**

Abteilung Bestandsentwicklung & Metadaten  
katharina.droese@sbb.spk-berlin.de

**Maria Federbusch**

Abteilung Handschriften und Historische Drucke  
maria.federbusch@sbb.spk-berlin.de

**Armin Möller**

Abteilung Informations- und Datenmanagement  
armin.moeller@sbb.spk-berlin.de

**Lole Westedt**

Abteilung Bestandsentwicklung & Metadaten  
lole.westedt@sbb.spk-berlin.de