

# Masterarbeit

zur Erlangung des akademischen Grades  
Master

**Technische Hochschule Wildau**  
**Fachbereich Wirtschaft, Informatik, Recht**  
**Studiengang Bibliotheks-informatik (M. Sc.)**

- Thema (deutsch):** Automatisierte Erwerbungsunterstützung – Konzeption und Implementierung eines Prototyps auf Basis bibliografischer Daten der Reihe O
- Thema (englisch):** Automated acquisition support – Conception and implementation of a prototype on the basis of bibliographic data of Series O

Autor/in: Nina Balz

Seminargruppe: BIM/17

Betreuer/in: M. Eng. Marcel-Dominique Paul Block

Zweitgutachter/in: Dipl.-Mathematiker Stefan Lohrum

Eingereicht am: 24.12.2019



## Abstract

Die Bayerische Staatsbibliothek, München, setzt in der Erwerbung deutschsprachiger Medien die Reihen A, B und M der Deutschen Nationalbibliografie<sup>1</sup> ein. Diese Reihen werden in elektronischer Form mithilfe eines Erwerbungsunterstützungsprogramms verarbeitet. Künftig soll auch die Reihe O (Online-Publikationen) in die Erwerbung einbezogen werden. Die vorliegende Arbeit untersucht, ob und wie das bestehende Programm oder ein anderes System am Markt für die Erwerbung auf der Basis von Reihe-O-Daten in der Bayerischen Staatsbibliothek angepasst werden kann. Vorgestellt wird ein Konzept für eine neue Anwendung, die im praktischen Teil der Arbeit prototypisch implementiert wird.

Schlagerworte: Bibliothek; Erwerbungsunterstützung; Automatisierung; Deutsche Nationalbibliografie; Reihe O

The Bavarian State Library, Munich, uses the Series A, B, and M of the German National Bibliography in its acquisition of German language media. These series are processed in electronic form with the help of an acquisition assistance program. In the future, Series O (Online Publications) shall also be included in the acquisition process. The present work examines whether and how the existing program or another system on the market can be adapted to the acquisition on the basis of Series O data in the Bavarian State Library. A concept for a new application is presented of which a prototype is implemented in the practical part of the work.

Keywords: Library; Acquisition Support; Automation; German National Bibliography; Series O

---

<sup>1</sup> S. Kap. 2.2.

## **Inhaltsverzeichnis**

<b>Abstract</b> .....	1
<b>Inhaltsverzeichnis</b> .....	2
<b>Abkürzungen</b> .....	4
<b>1. Einleitung</b> .....	5
1.1 Motivation.....	5
1.2 Zielstellung .....	5
<b>2. Ausgangssituation</b> .....	6
2.1 Geschäftsgänge der BSB in der deutschsprachigen Printerwerbung .....	6
2.2 Hintergrund .....	7
2.3 Arbeitsabläufe mit dem ABM-Tool.....	10
<b>3. Künftige Auswertung der Reihe O für die deutschsprachige Erwerbung</b> .....	13
3.1 Geplante Geschäftsgänge.....	13
3.2 Beurteilung der nötigen Veränderungen des ABM-Tools .....	15
3.3 Fazit.....	17
<b>4. Anforderungsanalyse</b> .....	17
4.1 Zweck.....	18
4.2 Allgemeine Beschreibung .....	18
4.3 Funktionale Anforderungen .....	19
4.4 Nicht-funktionale Anforderungen.....	21
<b>5. Marktanalyse</b> .....	22
5.1 Vorgehen.....	22
5.2 eMAS, Württembergische Landesbibliothek.....	23
5.3 Fazit.....	27
<b>6. Konzept und Prototyp</b> .....	27
6.1 Konzept.....	27
6.2 Softwareauswahl .....	38
6.3 Entwurfsmuster und Klassenkonzept.....	41
6.4 Weitere Dateien .....	49
6.5 Prototyp.....	50
6.6 Hinweise zur weiteren Programmierung .....	52
<b>7. Kritische Bewertung</b> .....	53
7.1 Laufzeit des Skripts.....	53
7.2 Speicherverbrauch.....	55

7.3 Mögliche Automatisierungsschritte.....	55
7.4 Ausgabe von Systemmeldungen .....	56
7.5 Implementierung des Controllers.....	56
7.6 Datenhaltung .....	56
7.7 Kodierung der Daten.....	57
<b>8. Ausblick .....</b>	<b>58</b>
8.1 Integration anderer nationalbibliografischer Reihen .....	58
8.2 Anpassung an die Bedürfnisse anderer Bibliotheken .....	59
<b>Anhang.....</b>	<b>60</b>
A 1 Glossar .....	60
A 2 Filterkriterien für Datensätze .....	61
A 3 Zuordnungskriterien für Datensätze der Pflichterwerbung zu einer zuständigen Bibliothek.....	62
A 4 Konkordanz Feldinhalt - XML-Tags – Attribute der Data-Klasse .....	62
A 5 Kodierung der Medientypen in Leader und Tag 008 .....	66
A 6 Exceptions .....	67
A 7 Datenbankschema: Tabelle titles .....	69
A 8 Datensätze und Dateigrößen .....	70
<b>Dokumentation.....</b>	<b>77</b>
D 1 Allgemeine Hinweise.....	77
D 2 Installation.....	77
D 3 Benutzung .....	78
D 4 PHP-Dokumentation .....	79
D 5 Tests .....	79
D 6 Inhalt der DVD.....	80
<b>Literaturverzeichnis .....</b>	<b>81</b>
<b>Abbildungsverzeichnis.....</b>	<b>84</b>
<b>Selbständigkeitserklärung.....</b>	<b>85</b>

## Abkürzungen

ABM-Tool	(An sich namenloses) BSB-internes Programm zur Erwerbungsunterstützung auf Basis der nationalbibliografischen Reihen A, B und M
AD	Amtliche Publikationen
bod-Titel	Im Book-on-Demand-Verfahren produzierter Titel. Von einer digitalen Vorlage wird erst bei Bestellung ein physisches Exemplar hergestellt.
BSB	Bayerische Staatsbibliothek, München
DDC-Sachgruppen	Fachliche Einteilung, die für alle Titel vergeben wird, die in den Reihen der Deutschen Nationalbibliografie angezeigt werden <sup>2</sup> .
DNB	Deutsche Nationalbibliothek, Frankfurt a. M. und Leipzig
GUI	Graphical User Interface, Benutzeroberfläche
RDA	Resource Description and Access; bibliothekarisches Regelwerk zur Erschließung von Medien
PflStG	Bayerisches Pflichtstückegesetz von 1986
WLB	Württembergische Landesbibliothek, Stuttgart

---

<sup>2</sup> Im einzelnen s. [https://www.dnb.de/DE/Professionell/DDC-Deutsch/DDCinDNB/ddcindb\\_node.html;jsessionid=8E80BEFF4615BF624103020119272C4F.internet282#doc259608bodyText1](https://www.dnb.de/DE/Professionell/DDC-Deutsch/DDCinDNB/ddcindb_node.html;jsessionid=8E80BEFF4615BF624103020119272C4F.internet282#doc259608bodyText1) (zuletzt gesehen 6.12.2019).

# 1. Einleitung

## 1.1 Motivation

Erwerbungs geschäftsgänge in deutschen Bibliotheken finden seit vielen Jahren rechnergestützt statt. Von der Auswahlentscheidung am Bildschirm über die Bestellung beim Lieferanten, die Einarbeitung der Rechnung, die (abschließende) Katalogisierung bis ggf. zur Bereitstellung der Medien - sofern es sich um elektronische handelt - werden meist mehrere Programme verwendet. Die Bemühungen richten sich ebenfalls seit langem darauf, diese Prozesse immer weiter zu integrieren, damit Daten möglichst nur noch einmal erfasst werden müssen und dann durch den Geschäftsgang hinweg immer wieder verwendet werden können.<sup>3</sup>

Zum Einsatz bei der Erwerbungsunterstützung kommen Programme unterschiedlicher Ausrichtung und Größenordnung: von Bibliothekssystemen internationaler Firmen über Plattformen von Bibliotheksverbänden und Medienlieferanten bis hin zu kleineren oder größeren Tools einzelner Bibliotheken, die für mitunter sehr spezifische Bedarfe geschrieben werden.

Auch die Bayerische Staatsbibliothek in München (BSB) setzt eine solche Eigenentwicklung für die Erwerbung von Medien aus Deutschland ein. Das ABM-Tool verarbeitet Daten der Reihen A, B und M der Deutschen Nationalbibliografie<sup>4</sup> für verschiedene Erwerbungsgebiete. Die Bibliothek strebt an, auch Daten der Reihe O aktiv und systematisch in der Erwerbung zu nutzen. Dafür ist das ABM-Tool in seiner jetzigen Form, u. a. wegen der im Vergleich zu den anderen Reihen erheblich größeren Datenmengen in der Reihe O, jedoch nicht geeignet.

## 1.2 Zielstellung

In der vorliegenden Arbeit werden anhand der Geschäftsgänge der BSB die Anforderungen an ein Programm formuliert, das Daten der Reihe O auswertet. In einer Marktanalyse wird untersucht, ob es Softwarelösungen gibt, die für die BSB übernommen und mit vertretbarem Aufwand für deren Bedürfnisse angepasst werden können. Es wird gezeigt, dass es ein solches Programm derzeit nicht gibt. Daher wird

---

<sup>3</sup> Zum Themenkomplex der Erwerbungsautomatisierung vgl. z. B. [Smith/Sasse, 2010].

<sup>4</sup> S. Kap. 2.2.

eine eigene Lösung konzipiert und im Praxisteil der Arbeit ein Prototyp programmiert, der die Umsetzung der Geschäftsgänge für die bayerische Pflichterwerbung in ihrer jetzigen Ausprägung ermöglicht.

## 2. Ausgangssituation

### 2.1 Geschäftsgänge der BSB in der deutschsprachigen Printerwerbung

Die deutschsprachige Printerwerbung der BSB teilt sich nach Erwerbungsarten in Kauf, Pflichtsammlung und Erwerbung amtlicher Publikationen. In allen drei Bereichen wird die Erwerbung – ausschließlich oder ergänzend – mit Daten der nationalbibliografischen Reihen A, B und M ausgeführt.

Die Käuferwerbung teilt sich auf in Erwerbung von Monografien, Zeitschriften und Noten, für die es jeweils eigene Referent\*innen<sup>5</sup> gibt. Die Monografienreferent\*innen betreuen jeweils bestimmte Fächer. Alle Referent\*innen wählen regelmäßig – Monografien i. d. R. wöchentlich, Noten monatlich, Zeitschriften in größeren Abständen – aus den Medien, die die Deutsche Nationalbibliothek (DNB) in den drei Reihen als Neuerwerbungen anzeigt, für die Erwerbung der BSB aus und leiten diese Titel an die Kaufteams, die sie bestellen.

Das Pflichtteam benutzt die Daten der drei Reihen zur Kontrolle, ob die bayerischen Verleger ihrer gesetzlichen Verpflichtung aus dem Pflichtstückegesetz vollständig nachgekommen sind.<sup>6</sup> Ggf. wird ein Mahnverfahren eingeleitet. Diese Kontrolle der neu angezeigten Titel gegen den Bestand der BSB wird zeitversetzt durchgeführt, um den Verlegern eine gewisse Zeit zur Ablieferung zu gewähren.

Amtliche Publikationen<sup>7</sup> können derzeit nicht in einem eigenen Geschäftsgang aus den Reihen A, B und M selektiert werden, da eine eigene Kennzeichnung amtlicher Publikationen in den nationalbibliografischen Reihen von der DNB vor Jahren eingestellt wurde. In der Erwerbungspraxis werden sie anhand ihres öffentlichen

---

<sup>5</sup> Es wird überwiegend geschlechtergerechte Sprache verwendet.

<sup>6</sup> S. Kap. 2.2.

<sup>7</sup> S. Kap. 2.2.

Herausgebers in den Kauf- und Pflicht-Geschäftsgängen identifiziert und an die Bearbeiter\*innen weitergeleitet.

### 2.2 Hintergrund

#### **Ablieferung von Pflichtexemplaren und amtlichen Publikationen**

Pflichtablieferung ist die gesetzlich geregelte Abgabe von Publikationen an Bibliotheken und andere staatliche Einrichtungen. Früher diente sie der Zensur, dem Urheberrechtsschutz oder der Gewährung von Privilegien. Heutzutage liegt in Deutschland ihr einziger Zweck in der Erhaltung des kulturellen Erbes einer Region oder des gesamten Bundesgebiets. Die DNB sammelt die Pflichtexemplare aus ganz Deutschland. Darüber hinaus gibt es in jedem Bundesland eine oder mehrere Bibliotheken, die für die Sammlung von Publikationen aus ihrer Region verantwortlich sind.<sup>8</sup>

Das bayerische Pflichtstückegesetz (PflStG) von 1986 bestimmt, dass von allen in Bayern veröffentlichten physischen Publikationen je zwei Exemplare an die BSB abzuliefern sind.<sup>9</sup> Diese archiviert ein Exemplar auf Dauer und gibt das zweite nach einem thematischen oder regionalen Schlüssel an eine von zehn anderen Bibliotheken in Bayern weiter.<sup>10</sup>

Derzeit (Stand: 12.2019) wird das Pflichtstückegesetz novelliert. Die beiden wichtigsten Änderungen der Novelle sind, die Abgabepflicht einerseits auf Netzpublikationen zu erweitern und sie andererseits auf nur noch ein Exemplar zu beschränken. Die Berechtigung, Netzpublikationen von Verlegern zu erhalten, wird jeweils bei einer bayerischen Bibliothek liegen, die in demselben Regierungsbezirk ihren Sitz hat wie der abgebende Verleger. Die organisatorische Abwicklung der Pflichtsammlung im Kontakt mit den Verlegern wird wie bisher bei der BSB angesiedelt sein.

Es ist geplant, die Sammlung von Netzpublikationen schrittweise aufzunehmen. So soll mit E-Books ausgewählter Verlage im PDF-Format begonnen und der Kreis der Verlage sowie später der Publikationsarten und Formate allmählich ausgeweitet werden.

---

<sup>8</sup> Vgl. [Strauch/Rehm, 2007], 'Pflichtablieferung', S. 342; 'Pflichtexemplar', S. 342-343.

<sup>9</sup> S. [PflStG, 1986], Art. 1 Abs. 1 Satz 1.

<sup>10</sup> Vgl. [Pflichtstückesammlung, 1986], S. 530.

Eine Untergruppe von Pflichtexemplaren stellen amtliche Publikationen dar, also Schriften, die von Behörden, Dienststellen und Einrichtungen des öffentlichen Rechts herausgegeben werden. Sie sind in Deutschland seit der zweiten Hälfte des 19. Jahrhunderts aus den Pflichtstück- und Pressegesetzen ausgenommen und unterliegen eigenen rechtlichen Regelungen.<sup>11</sup> Sie werden ebenfalls auf Bundes- und auf Landesebene gesammelt. Die BSB ist als eine von vier Bibliotheken in Deutschland nicht nur für die Sammlung amtlicher Publikationen ihres eigenen Bundeslands zuständig, sondern auch für die Sammlung von Publikationen aus ganz Deutschland.<sup>12</sup>

Seit 2009 besteht eine Ablieferungspflicht auch für bayerische amtliche Publikationen in elektronischer Form.<sup>13</sup> Allerdings werden aus technischen Gründen noch nicht alle Veröffentlichungen aller bayerischen Herausgeber gesammelt, ebenso bei weitem noch nicht alle Publikationen aus weiteren Bundesländern mit elektronischer Ablieferungspflicht. Die Ablieferung wird derzeit intellektuell kontrolliert, indem die Websites abliefernder Institutionen überprüft werden.

### **Die Reihe O der Deutschen Nationalbibliografie**

Die Deutsche Nationalbibliografie wird von der DNB herausgegeben und erscheint seit 1913. Sie verzeichnet alle Erwerbungen der DNB; dabei handelt es sich um „alle Werke in Schrift und Ton, die seit 1913 in und über Deutschland oder in deutscher Sprache veröffentlicht werden“<sup>14</sup>. Sie erscheint in verschiedenen Reihen, die Publikationen nach unterschiedlichen inhaltlichen oder formalen Kriterien zusammenfassen und wochen- oder monatsweise publiziert werden. Die Reihe A enthält Bücher, Zeitschriften, nicht-musikalische Tonträger, weitere audiovisuelle Medien, Mikroformen und elektronische Publikationen des Verlagsbuchhandels, die Reihe B dieselben

---

<sup>11</sup> Vgl. [Strauch/Rehm, 2007], 'Amtliche Druckschriften', S. 11-12. – Hier wird die Zahl der bundesweit sammelnden Bibliotheken fälschlicherweise mit drei angegeben. Es fehlt die Staatsbibliothek Berlin – Preußischer Kulturbesitz (vgl. <https://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandsaufbau/amtsdruckschriften/abgabeerlasse/>; zuletzt gesehen 6.12.2019).

<sup>12</sup> S. <https://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandsaufbau/amtsdruckschriften/abgabeerlasse/>; <https://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandsaufbau/amtsdruckschriften/deutsche-amtsdruckschriften/> (zuletzt gesehen 6.12.2019).

<sup>13</sup> S. [Abg-Bibl, 2009].

<sup>14</sup> Die Deutsche Nationalbibliothek im Porträt ([https://www.dnb.de/DE/Ueber-uns/Portraet/portraet\\_node.html#doc57856bodyText1](https://www.dnb.de/DE/Ueber-uns/Portraet/portraet_node.html#doc57856bodyText1)) (zuletzt gesehen 6.12.2019).

Medienarten, sofern sie außerhalb des Verlagsbuchhandels erscheinen. Die Reihe M verzeichnet Musikalien (Noten).<sup>15</sup> Daneben gibt es weitere Reihen.

Die Reihe O, die Onlinepublikationen enthält, erscheint seit 2010.<sup>16</sup> Sie wurde aus den anderen nationalbibliografischen Reihen ausgegliedert, als die Menge an Online-Publikationen stark zunahm. Verzeichnet werden in ihr derzeit E-Books, E-Zeitschriften, E-Paper, Noten in elektronischer Form, archivierte Websites sowie Hörbücher und Filme, die trägerlos erscheinen.<sup>17</sup>

Publikationen, die parallel in einer gedruckten Auflage wie auch in elektronischer Form erscheinen, werden sowohl in der Reihe O als auch in einer der anderen nationalbibliografischen Reihen nachgewiesen. Hingegen werden Publikationen, die parallel in elektronischer Form wie auch gedruckt, jedoch im Book-on-Demand-Verfahren angeboten werden, primär in der Reihe O angezeigt. Grund hierfür ist die Sammelpraxis der DNB: Solche Publikationen werden zunächst nur in elektronischer Form gesammelt und in der Reihe O nachgewiesen. Erst wenn der Verlag 25 gedruckte Exemplare verkauft hat, werden auch zwei gedruckte Exemplare angefordert und in der entsprechenden Reihe angezeigt. Diese Besonderheit ist für die bayerische Pflichtablieferung von Bedeutung.<sup>18</sup>

Der Zugang in der Reihe O entwickelte sich von 166.000 Einheiten im Einführungsjahr 2010 über 612.000 Einheiten fünf Jahre später bis zuletzt knapp 1,4 Mio. Einheiten im Jahr 2018. Der Bestand belief sich auf über 5,8 Mio. Einheiten zu Ende 2018.<sup>19</sup> Die weitere Entwicklung wird von verschiedenen Faktoren wie Veränderungen in der Verlagslandschaft und dem künftigen Publikationsaufkommen abhängen und ist insofern kaum näher zu prognostizieren. Zu erwarten ist jedoch in jedem Fall eine weitere Steigerung.<sup>20</sup>

---

<sup>15</sup> S. [https://www.dnb.de/DE/Professionell/Metadatendienste/Metadaten/Nationalbibliografie/nationalbibliografie\\_node.html#doc58256bodyText2](https://www.dnb.de/DE/Professionell/Metadatendienste/Metadaten/Nationalbibliografie/nationalbibliografie_node.html#doc58256bodyText2) (zuletzt gesehen 6.12.2019).

<sup>16</sup> S. Datendienst Deutsche Nationalbibliografie: Reihencode und Kennzeichnung von Print-on-demand-Publikationen der Reihe O. Online verfügbar unter [https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Metadatendienste/Rundschreiben/rundschreiben20100131\\_reihencodePrintOnDemand.pdf?\\_\\_blob=publicationFile&v=2](https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Metadatendienste/Rundschreiben/rundschreiben20100131_reihencodePrintOnDemand.pdf?__blob=publicationFile&v=2) (zuletzt gesehen 6.12.2019).

<sup>17</sup> Entsprechend Auswertungen der Verfasserin anhand von Reihe-O-Titeln.

<sup>18</sup> S. Kap. 6.5.

<sup>19</sup> S. Jahresberichte der Deutschen Nationalbibliothek, <http://d-nb.de/jahresbericht> (zuletzt gesehen 6.12.2019).

<sup>20</sup> So auch formuliert in E-Mail-Auskünften der DNB vom 13.9. und 16.9.2019.

Pro Arbeitstag<sup>21</sup> wurden in den ersten drei Quartalen des Jahres 2019 im Durchschnitt etwa 2.500 Datensätze in der Reihe O angezeigt. Allerdings schwankt die Anzahl zwischen 153 (2.6.2019) und 23.889 (27.2.2019) stark. In früheren Jahren wurden bei Stichproben noch höhere Zahlen gefunden (43.322 Sätze für den 1.12.2017). Ursache hierfür ist, dass die vielen großen und kleinen deutschen Verlage an die DNB blockweise und nach ihrem im Detail selbstgewählten Zeitplan abliefern, so dass sich in der Summe diese Unregelmäßigkeiten ergeben.<sup>22</sup>

Die Daten der Reihe O können über den OPAC der DNB abgefragt oder z. B. über eine OAI-Schnittstelle in den Formaten MARC21-XML, DNB Casual und RDF abgerufen werden.<sup>23</sup> Daneben gibt es eine Reihe anderer Bezugsmöglichkeiten.<sup>24</sup>

### 2.3 Arbeitsabläufe mit dem ABM-Tool

Das für die Bearbeitung der Reihen A, B und M eingesetzte Programm ist eine (eigentlich namenlose) Hausentwicklung der BSB.<sup>25</sup> Es ist in PHP programmiert und setzt eine relationale Datenbank (MariaDB) ein, die mit SQL angesprochen wird. Für Benutzerinteraktionen gibt es eine Oberfläche. Alle Referent\*innen der deutschsprachigen Erwerbung sowie ein Mitarbeiter der Pflichtstelle arbeiten mit dem ABM-Tool.

Die Referent\*innen rufen auf der Benutzeroberfläche mithilfe eines Suchstrings die Medien einer Wochen- oder Monatslieferung der DNB aus deren Online-Katalog ab. Die Suchstrings sind in Contextual Query Language (CQL) formuliert<sup>26</sup> und kombi-

---

<sup>21</sup> Montag bis Samstag.

<sup>22</sup> S. Auswertung in Anhang A 8 Datensätze und Dateigrößen.

<sup>23</sup> S. <https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai.html?nn=58282#doc58284bodyText9> (zuletzt gesehen 6.12.2019). Vgl. a. Kap. 6.1.

MARC21-XML ist ein xml-basiertes internationales Austauschformat für bibliothekarische Metadaten (s. [https://www.dnb.de/DE/Professionell/Metadatendienste/Exportformate/MARC21/marc21\\_node.html#doc158972bodyText1](https://www.dnb.de/DE/Professionell/Metadatendienste/Exportformate/MARC21/marc21_node.html#doc158972bodyText1); zuletzt gesehen 6.12.2019).

DNB Casual ist ein xml-basiertes Dublin-Core-Format, das gegenüber anderen Exportformaten nur ein reduziertes Elementeset enthält (s. <https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai.html?nn=58282#doc58284bodyText9>; zuletzt gesehen 6.12.2019).

RDF (Resource Description Framework) ist ein Austauschformat für das semantische Web (s. <https://www.w3.org/RDF/>; zuletzt gesehen 6.12.2019).

<sup>24</sup> S. [https://www.dnb.de/DE/Professionell/Metadatendienste/metadatendienste\\_node.html#sprg186916](https://www.dnb.de/DE/Professionell/Metadatendienste/metadatendienste_node.html#sprg186916) (zuletzt gesehen 6.12.2019)

<sup>25</sup> Zum ABM-Tool gibt es keine Veröffentlichungen. Es ist nur im BSB-Intranet aufzurufen.

<sup>26</sup> S. <https://www.dnb.de/DE/Service/Hilfe/Katalog/kataloghilfeExpertensuche.html> (zuletzt gesehen 6.12.2019).

nieren positive Kriterien und Ausschlusskriterien (Zeitraum, Medienart usw.).

(S. Abb. 1.)

### **CQL-Suchstrings**

Beispiel für Kauf:

```
(hsg=330 OR hsg=380 OR hsg=650) AND (wvn=13,A01*) NOT vfo=XA-DE-BY  
NOT bbg=Ab* NOT mat=discs NOT cod=tt NOT (cod=vi OR cod=vh OR cod=vb  
OR sw="dvd-video") NOT cod=lo
```

Beispiel für Pflicht:

```
(hsg=0** OR hsg=1** OR hsg=2** OR hsg=3** OR hsg=4** OR hsg=5** OR  
hsg=6** OR hsg=7** OR hsg=8** OR hsg=9** OR hsg=S OR hsg=K) AND  
(wvn=13,A01* OR wvn=13,B01*) AND vfo=XA-DE-BY
```

Kodiert sind die DDC-Hauptsachgruppe (hsg), Jahr, Reihe und Wochenlieferung (wvn), Länderkenzeichen (vfo) sowie verschiedene Materialarten (mat, cod).

*Abb. 1. Suchstrings des ABM-Tools in CQL-Syntax*

Das Skript sendet diese Suchanfrage an die Druck-Schnittstelle des DNB-Katalogs, der als Ergebnis eine HTML-Seite mit maximal 500 Treffern zurücksendet. Das Skript parst diese HTML-Daten und erzeugt eine gefelderte Darstellung der Titeldatensätze im HTML-Format. (S. Abb. 2.)

Die Referent\*innen prüfen die Datensätze der Reihe nach auf ihre Erwerbungsrelevanz. Ihre Erwerbungsentscheidung können sie in einem Freitextfeld unter jedem Datensatz vermerken.

Für alle Datensätze kann ein Bestandsabgleich gegen den Online-Katalog (OPAC) der BSB ausgeführt<sup>27</sup> und das Ergebnis im Datensatz angezeigt werden. Bereits vorhandene Titel können so bei der Erwerbung übergangen werden.

Die Referent\*innen können auf der Benutzeroberfläche zudem Filter einstellen, die Datensätze nach weiteren Kriterien mit Standardangaben im Freitextfeld kennzeichnen, z. B. Titel, die außerhalb Deutschlands erschienen sind oder unter 100 Seiten

<sup>27</sup> Zur BSB-Query-Schnittstelle s. <https://www.bsb-muenchen.de/recherche-und-service/suchen-und-finden/direktlinks-fuer-suchanfragen-kurzanleitung-bsb-query-schnittstelle/> (zuletzt gesehen 6.12.2019).

## 2. Ausgangssituation

umfassen. Dadurch wird die Erwerbungs Auswahl weiter beschleunigt. Diese Filter greifen auf TXT-Dateien zurück, die im Intranet hinterlegt sind und deren Inhalt von den beiden bibliothekarischen Betreuer\*innen gepflegt werden kann.

The screenshot displays the ABM-Tools interface. At the top, the browser address bar shows the URL: <https://intranet.bsb-muenchen.de/BSB-DNB-Lieferung/index.php?fmode=Ung>. The page header includes the BSB Bayerische Staatsbibliothek logo and the title 'KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK'. A search bar contains the query: `(hsg=330 OR hsg=380 OR hsg=650) AND (wvn=19,A21*) NOT geo=XA-DE-BY NOT bbg=A`. Below the search bar is a grid of selection options with checkboxes and labels such as 'BSB-Katalog', 'AD', 'Lfd. e Zs.', 'E-Paket - E-Book oder Print?', 'Lang - Not - Auswahl über Verlagsliste, falls laufendes EJahr', 'Rot', 'Leihmat.', 'kholzle', 'Bern. suchen', 'PV', 'AD Verlag', 'AT', 'E-Ress.', 'Gruyter - E-Book oder Print?', 'Gelb', 'nbalz', 'fgoetz', 'Kauf', 'Pflicht', 'FF', 'G/T', 'CH', 'MT91?', 'Wiley - E-Book wenn für Lesesaal', 'Not', 'khundham', and 'Markieren x'. Below the grid are sections for 'Bemerkungen:' (Speichern, Löschen), 'Filtern:' (0, Ungefiltert, Gefiltert, Erledigt), and 'Export:' (Kauf, Pflicht). The search results section shows 'Ungefiltert (0/111)', 'Gefiltert (0/0)', and 'Erledigt (0/0)'. The search criteria are repeated: `Suchergebnis - Suche nach: (hsg=330 or hsg=380 or hsg=650) and (wvn=19,A21*) not geo=XA-DE-BY not bbg=Ab* not mat=discs not cod=tt not (cod=vi or cod=vh or cod=vb or sw="dvd-video") not cod=lo, Treffer 1-111 von insgesamt 111 - 19.11.19 09:30`. Two results are shown: result 37 for 'Finanz- und Rechnungswesen; Teil: Band 3., Vertiefungen' by Brodmann, Daniel, and result 84 for 'BankCollee: Teil: 3. Semester. / Betriebswirtschaft'.

Abb. 2. Die Oberfläche des ABM-Tools

Die Datensätze werden mit dem Inhalt des Freitextfelds per Buttonklick in der Datenbank gespeichert. Damit kann die Arbeit an einer Lieferung auch unterbrochen und zu einem späteren Zeitpunkt fortgesetzt werden. Nach Abschluss des Erwerbungsprozesses werden markierte Datensätze aus der Datenbank abgerufen, in einer RTF-Datei gespeichert und per E-Mail an ein Erwerbungs team oder eine\*n Kollege\*in versandt. Als Selektionskriterium dienen Textanfänge von Einträgen in das Freitextfeld.

In ähnlicher Weise rufen Mitarbeiter\*innen der Pflichtstelle per CQL-Suchstring pflichtrelevante Titel auf<sup>28</sup> und stoßen einen Bestandsabgleich an. Noch nicht vorhandene Titel werden durch das ABM-Tool auf verschiedene Regierungsbezirke aufgeteilt; der Zuordnung dienen Konkordanzen zwischen bayerischen Orten und den Regierungsbezirken, in denen die Orte liegen. Diese Titel werden ebenfalls per E-Mail an die Bearbeiter\*innen versandt, die jeweils für diese Regierungsbezirke zuständig sind.

Amtliche Publikationen können nicht bereits beim Datenabruf durch einen CQL-Suchstring separat abgerufen werden, da sie in den nationalbibliografischen Reihen nicht mehr gekennzeichnet werden. Sie werden daher durch eine Filterung nach amtlichen Herausgebern in den Kauf- und Pflicht-Geschäftsgängen identifiziert und an die zuständigen Bearbeiter\*innen weitergeleitet.

## **3. Künftige Auswertung der Reihe O für die deutschsprachige Erwerbung**

### **3.1 Geplante Geschäftsgänge**

Künftig soll auch die Reihe O in die Geschäftsgänge der deutschsprachigen Erwerbung an der BSB einbezogen werden. Das betrifft folgende Bereiche:

#### Kaufwerbung

– In einigen, noch nicht festgelegten Fächern sollen von den Referent\*innen Monografien in elektronischer Form ausgewählt werden. Dabei soll auf kostenpflichtige Publikationen im PDF-Format eingeschränkt werden. Parallel in Print und in elektronischer Form erscheinende Publikationen werden, sofern sie in einer regulären Auflage erscheinen, auch in den Reihen A und ggf. B oder M angezeigt, die ohnehin von den Referent\*innen durchgesehen werden. Daher sind diese Titel in der Reihe O wegzufiltern. Angezeigt werden sollen hingegen Titel, die parallel im bod-Verfahren erscheinen.

---

<sup>28</sup> Gefiltert über das in einem Teil der Daten vorhandene Länderkennzeichen (für Bayern: XA-DE-BY).

### 3. Künftige Auswertung der Reihe O für die deutschsprachige Erwerbung

---

- Neu anlaufende Periodika sollen ohne weitere Einschränkungen angezeigt werden und von der zuständigen Referentin gesichtet werden können.
- Noten sollen der Referentin angezeigt werden, wenn sie aus bestimmten Verlagen stammen.

Weitere Kriterien wie das vom Verlag angebotene Erwerbungsmodell (Kauf vs. Lizenz) oder das Vorhandensein eines DRM (digitales RechteManagement) können nicht aus der Reihe O ermittelt werden und müssen nachgelagert von den Erwerbungs specialist\*innen intellektuell abgeprüft werden.

#### Pflichterwerbung

- Hier soll eine Lücke in der Ablieferungskontrolle geschlossen werden: Denn Publikationen, die im bod-Verfahren produziert werden und sowohl in gedruckter als auch in elektronischer Form erscheinen, werden von der DNB primär in der Reihe O nachgewiesen.<sup>29</sup> Diese Titel können derzeit von der BSB angesichts der Menge der angezeigten Publikationen in der Reihe O von den Mitarbeiter\*innen der Pflichtstelle nicht manuell überprüft werden. Derzeit handelt es sich ausschließlich um Publikationen des Münchner Verlags GRIN.
- Nach Inkrafttreten der Novelle des bayerischen Pflichtstückegesetzes<sup>30</sup> wird ein Programm benötigt, das für die Netzpublikationen bayerischer Verlage die Ablieferungskontrolle ausführt. Im Unterschied zum derzeitigen Geschäftsgang ist die Bestandsprüfung nicht mehr in jedem Fall gegen den Bestand der BSB, sondern dann gegen den Bestand der jeweils berechtigten Pflichtbibliothek auszuführen.

#### Erwerbung amtlicher Publikationen

- Es soll ermöglicht werden, zur Kontrolle der Ablieferung amtlicher Publikationen sowohl aus Bayern als auch aus anderen Bundesländern Titelsätze von Publikationen ablieferpflichtiger Institutionen aus der Reihe O zu filtern und sie gegen den Bestand der BSB abzuprüfen. Nicht vorhandene Publikationen sollen nach Bundesländern sowie bayerische Publikationen nach Regierungsbezirken aufgeteilt und den zuständigen Bearbeiter\*innen zur Verfügung gestellt werden.

Bei der Filterung der Datensätze der Reihe O werden folgende Kriterien benötigt:

---

<sup>29</sup> S. Kap. 2.2.

<sup>30</sup> S. Kap. 2.2.

- Medientyp (Monografie, Zeitschrift, unselbständiges Werk wie z. B. Aufsatz usw.)
- DDC-Sachgruppen
- Dateiformat (PDF, epub usw.)
- amtlicher Herausgeber
- Erscheinen eines Titels parallel in elektronischer und körperlicher Form (Kauf, Pflicht)
- Herstellung eines Titels im bod-Verfahren
- Kostenpflichtigkeit
- Verlagsnamen bayerischer Verlage
- ISBN-Stämme bayerischer Verlage
- Liste bayerischer Erscheinungsorte

#### **3.2 Beurteilung der nötigen Veränderungen des ABM-Tools**

Das ABM-Tool in seiner jetzigen Form kann Datensätze der Reihe O ebenso verarbeiten wie die aller anderen nationalbibliografischen Reihen, da alle Datensätze von der DNB einheitlich nach dem bibliothekarischen Regelwerk RDA erschlossen werden und über dieselben Schnittstellen der DNB zugänglich sind.

Allerdings treten bei der Bearbeitung einige gewichtige Probleme auf, die z. T. spezifisch für die Bearbeitung der Reihe O sind, z. T. auch schon bei den anderen Reihen auftreten, aber hier an Schärfe gewinnen:

- Bei dem in Kap. 2.3 beschriebenen Datenabruf über die Druck-Schnittstelle der DNB ist die gelieferte Trefferzahl anbieterseitig auf maximal 500 begrenzt. Diese Zahl kann nicht weiter erhöht werden. Die Mengenbeschränkung stellt bereits für die Bearbeitung der Reihen A, B und M ein Problem dar.<sup>31</sup> Für diese Reihen wurde ein Workaround implementiert, in dem Referent\*innen, deren Suchanfragen regelmäßig mehr als 500 Treffer zurückliefern, diese in mehrere Anfragen aufteilen. Das bedeutet bereits jetzt einen größeren Aufwand.

---

<sup>31</sup> In der Entwicklung des ABM-Tools war die ursprünglich geplante Datenabfrage über die SRU-Schnittstelle noch nicht sinnvoll realisierbar, weil über SRU nicht alle für die Erwerbung benötigten Feldinhalte ausgeliefert wurden. Über die HTML-Trefferanzeige standen sie jedoch zur Verfügung. – Inzwischen liefert auch die SRU-Schnittstelle alle diese Feldinhalte aus. – Zur SRU-Schnittstelle s. a. Kap. 6.1.

### 3. Künftige Auswertung der Reihe O für die deutschsprachige Erwerbung

---

Die Reihe O liefert erheblich mehr Daten als die anderen Reihen, mit stetig steigender Tendenz.<sup>32</sup> Damit würde im Laufe der Zeit eine immer feinere Aufteilung der Suchanfragen nötig werden und der Aufwand immer weiter steigen.

– Problematisch ist in diesem Zusammenhang auch, dass der Datenabruf über viele einzeln konstruierte, komplexe Suchstrings organisiert ist, anstatt dass die Datenmenge eines Zeitraums in ihrer Gesamtheit abgerufen und auf die einzelnen Erwerbungsgebiete verteilt wird. Durch die vielen Suchstrings, die von einigen Referent\*innen auch eigenständig angepasst werden, besteht die Gefahr, dass erwerbungsrelevante Daten versehentlich mehrfach oder aber gar nicht abgerufen werden.

– Für die Auswahl elektronischer Publikationen werden zusätzliche Angaben benötigt, z. B. die Kosten- oder Lizenzpflicht oder das Erscheinen als elektronische und parallel als bod-Publikationen. Nicht alle Felder, in denen diese Angaben stehen, sind für die CQL-Suche indiziert. Auch werden über die Druck-Schnittstelle nicht alle Feldinhalte ausgegeben, sondern nur eine Auswahl. Parallel erscheinende bod-Publikationen sind dadurch gar nicht zuverlässig zu identifizieren. Sie spielen jedoch für die bayerische Pflichterwerbung sowohl vor als auch nach der geplanten Gesetzesnovellierung eine wichtige Rolle.<sup>33</sup>

Die Umstellung des Datenabrufs von der Druck- auf die SRU- oder OAI<sup>34</sup>-Schnittstelle wäre möglich, würde dann jedoch erfordern, das Parsen von HTML- auf XML-Dateien umzustellen.

– Derzeit ist für alle Titel eine Bestandsprüfung implementiert, die Titel anhand ihrer ISBN oder ISSN mit dem BSB-OPAC abgleicht. Für Netzpublikationen, die dem novellierten bayerischen Pflichtstückegesetz unterliegen werden, müsste dieser Abgleich so verändert werden, dass die Titel mit dem Bestand derjenigen Bibliothek abgeglichen werden, die als Pflichtbibliothek für sie zuständig ist.

Dazu müsste zunächst zwischen Kauf- und Pflichttiteln unterschieden werden. Für Pflichttitel müsste die Zuordnung zu einer Pflichtbibliothek von einer Kombination

---

<sup>32</sup> S. Kap. 2.2. Die Deutsche Nationalbibliografie. Vgl. die Zugänge im physischen und unkörperlichen Bereich in den Jahresberichten der DNB ([www.dnb.de/jahresbericht](http://www.dnb.de/jahresbericht), zuletzt gesehen 6.12.2019).

<sup>33</sup> S. Kap. 2.2, 3.1.

<sup>34</sup> Zur SRU-Schnittstelle (Search/Retrieve via URL) s. [https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/SRU/sru\\_node.html](https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/SRU/sru_node.html) (zuletzt gesehen 6.12.2019). Zur OAI-Schnittstelle s. Kap. 6.1.

mehrerer Kriterien abhängig gemacht werden.<sup>35</sup> Diese differenzierte Zuordnung ist über das System einfacher Filterdateien nicht abzubilden. Auch hier müsste eine tiefere Veränderung des Skripts erfolgen.<sup>36</sup>

Insgesamt betrachtet, sind einige größere Anpassungen vorzunehmen. Bei dem ABM-Tool handelt es sich um eine Anwendung mit nur wenigen Elementen objektorientierter Programmierung. Aufgrund dessen ist anzunehmen, dass die nötigen Anpassungen nur mit größerem Aufwand umzusetzen wären. Es wäre daher sinnvoller, die Anwendung nach dem Paradigma der Objektorientierung vollständig neu aufzubauen.

### 3.3 Fazit

Aus den in Kap. 3.2 genannten Gründen wird die Entscheidung getroffen, anstelle einer Anpassung des bestehenden Tools nach einem anderen Programm zu suchen, das leichter anzupassen ist, oder eine eigene Anwendung zu konzipieren.

Daher wird im nächsten Schritt eine Anforderungsanalyse für ein Programm erstellt, das die von der BSB geplanten Erwerbungs geschäftsgänge mit Reihe-O-Daten ermöglicht.

## 4. Anforderungsanalyse

Vorbemerkung: Die Anforderungsanalyse ist angelehnt an die IEEE Recommended Practice for Software Requirements Specifications (SRS IEEE 830-1998). Für die Darstellung wird eine Vorlage von Karl. E. Wiegers<sup>37</sup> verwendet.

Hingewiesen wird auf das Glossar in Anhang A 1. Im Glossar erklärte <Begriffe> sind im gesamten Kapitel 4 mit spitzen Klammern gekennzeichnet.

---

<sup>35</sup> S. Anhang A 3 Zuordnungskriterien für Datensätze der Pflichterwerbung zu einer zuständigen Bibliothek.

<sup>36</sup> Zudem bestehen für die praktische Arbeit mit den Reihen A, B und M weitere kleinere Desiderate, auf die hier nicht näher eingegangen werden soll.

<sup>37</sup> Vgl. [SRS IEEE 830-1998], [Wiegers, 1990]).

#### 4.1 Zweck

Das Programm stellt in regelmäßigen Abständen Daten der Reihe O (Online-Publikationen) der Deutschen Nationalbibliografie für Mitarbeitende der Bayerischen Staatsbibliothek zur Verfügung. Die Daten werden so aufbereitet, dass sie als Auswahlgrundlage in der Medien-Kaufwerbung und zur Ablieferungskontrolle in der Erwerbung von Pflicht- und amtlichen Publikationen dienen können.

#### 4.2 Allgemeine Beschreibung

##### Benutzerbeschreibung

<b>Benutzergruppe</b>	<b>IT-Erfahrung</b>	<b>Benutzerinteresse</b>	<b>Frequenz der Systemnutzung</b>
Kaufreferent*innen der deutschsprachigen Erwerbung (ca. 8 Personen)	Internet, bibliothekarische Datenbanken	Kaufauswahl aufgrund von Datensätzen, die auf einer Benutzeroberfläche angezeigt werden	wöchentlich
Kaufteams (2–4 Teams)	Internet, bibliothekarische Datenbanken	Bestellung der von den Kaufreferent*innen ausgewählten Medien	täglich
Mitarbeitende der Pflichtstelle (ca. 8 Personen)	Internet, bibliothekarische Datenbanken	Reklamation von Medien, die in Bayern erschienen sind und noch nicht vom jeweiligen Verlag an die BSB abgeliefert wurden	monatlich
Mitarbeitende für amtliche Publikationen (ca. 8 Personen)	Internet, bibliothekarische Datenbanken	Reklamation von Medien, die von amtlichen Herausgebern veröffentlicht wurden und noch nicht an die BSB abgeliefert wurden	monatlich
Ggf. zentrale*r Bearbeiter*in aus einer der o. g. Benutzergruppen	Internet, bibliothekarische Datenbanken, Pflegeerfahrung in anderen bibliothekarischen Systemen	Pflege von im System hinterlegten Informationen, die häufigen Änderungen unterliegen (z. B. Listen bayerischer Verlage)	nach Bedarf, mindestens wöchentlich

**Betriebsumgebung**

Betriebssystem: NN<sup>38</sup>

Browser: Firefox ESR und/oder Internet Explorer (jeweils neueste Version)

Textverarbeitung: Windows Office (7, 10), Open Office

**Software (bei Eigenentwicklung oder eigener Weiterentwicklung)**

<b>ID</b>	<b>Beschreibung</b>	<b>Gewichtung</b>
SW-REQ-1	Die eingesetzte Software (Programmiersprachen u. a.) wird von Mitarbeitenden des IT-Referats der BSB beherrscht.	muss
SW-REQ-2	Die eingesetzte Software ist weit verbreitet.	muss
SW-REQ-3	Die eingesetzte Software ist an der BSB vorhanden oder Freeware.	muss

**4.3 Funktionale Anforderungen**

<b>ID</b>	<b>Beschreibung</b>	<b>Gewichtung</b>
REQ-1	Aktuelle Datensätze der Reihe O werden mindestens wochenweise ins System importiert.	muss
REQ-2	Die Datenfelder der Datensätze werden gefiltert. (S. Anhang A 5.)	muss
REQ-3	Eine Filterung der Datenfelder ist für die Erwerbung von Medien nach unterschiedlichen Kriterien für die verschiedenen <Erwerbungsarten> möglich.	muss
REQ-4	Die Filterung der Datenfelder ist veränderbar.	muss
REQ-5	Datensätze werden gefiltert.	muss
REQ-6	Die Filterung der Datensätze unterscheidet sich nach den <Erwerbungsarten>. (S. Tabelle A 2).	muss

<sup>38</sup> Von dem IT-Referat der BSB nicht mitgeteilt.

<b>ID</b>	<b>Beschreibung</b>	<b>Gewichtung</b>
REQ-7	Filterkriterien für Datensätze sind als Negativfilter (NOT X) einsetzbar.	muss
REQ-8	Die gefilterten Datensätze werden einer <zuständigen Bibliothek> zugeordnet. (S. Tabelle A 3.)	muss
REQ-9	Die gefilterten Datensätze werden gegen den Bestand der <zuständigen Bibliothek> überprüft.	muss
REQ-10	Die Bestandsprüfung kann für die einzelnen <Erwerbungsarten> zu unterschiedlichen Zeitpunkten vorgenommen werden.	muss
REQ-11	Die GUI für die <Kaufreferent*innen> zeigt die modifizierten Datensätze an.	muss
REQ-12	Die Datensätze werden in strukturierter Form angezeigt.	muss
REQ-13	Alle <Kaufreferent*innen> tragen auf der GUI bei jedem Datensatz ihre <Kaufentscheidungen> ein.	muss
REQ-14	Die <Kaufreferent*innen> können den Auswahlprozess unterbrechen und zu einem späteren Zeitpunkt fortsetzen.	muss
REQ-15	Die <Kaufreferent*innen> weisen Datensätze, die in die Zuständigkeit eines anderen <Kaufreferenten> fallen, diesem zu.	soll
REQ-16	Datensätze, die zum Kauf ausgewählt sind, werden einem der <Kaufteams> zugeordnet.	muss
REQ-17	(bei Eigenentwicklung <sup>39</sup> ;) Datensätze des Münchener Verlags GRIN werden nach vom Verlag vergebenen Medientypen gefiltert.	muss
REQ-18	Datensätze, die für die Pflichterwerbung gefiltert worden sind und im Bestand der <zuständigen Bibliothek> nicht vorhanden sind, werden Mitarbeitenden der <Pflichtstelle> zur Bearbeitung zugewiesen.	muss
REQ-19	Datensätze, die als <amtliche Publikationen> gefiltert worden sind und im Bestand der BSB nicht	muss

<sup>39</sup> Es handelt sich um eine sehr spezielle Anforderung, die von keinem System eines externen Herstellers oder einer anderen Bibliothek erwartet werden kann.

ID	Beschreibung	Gewichtung
	vorhanden sind, werden Mitarbeitenden des <Unberechneten Zugangs> zugewiesen.	

#### 4.4 Nicht-funktionale Anforderungen

##### Qualität

NF-REQ-1	Das System ist einfach auf die Betriebsumgebung der BSB zu portieren.	muss
NF-REQ-2	Das System ist einfach zu warten.	muss
NF-REQ-3	Das System ist nach der Novellierung des bayerischen Pflichtstückegesetzes an die veränderten Filterkriterien anzupassen.	muss
NF-REQ-4	Das System ist für die Erwerbung auf Basis der Reihen A, B und M einzusetzen.	kann

##### Kosten

NF-REQ-5	<b>(bei Lizenzierung:)</b> einmalig im unteren vierstelligen Eurobereich für Lizenzierung und Anpassung	muss
----------	---	------

##### Daten

NF-REQ-6	Alle verwendeten Daten, insbesondere die Reihe O, sollen kostenfrei und frei zugänglich sein (ggf. nach einer Anmeldung).	muss
----------	---	------

Im folgenden Schritt wird untersucht, ob bereits in anderen Bibliotheken Softwarelösungen eingesetzt werden, die für die BSB angepasst und übernommen werden können.

## 5. Marktanalyse

### 5.1 Vorgehen

Bei der Marktanalyse sollen vorhandene Programme ermittelt und untersucht werden, die in Bibliotheken zur Bearbeitung der Reihe O in der Erwerbung eingesetzt werden oder sich grundsätzlich, ggf. nach überschaubaren Anpassungen, dafür einsetzen lassen. Dabei sollen in der Erwerbung Geschäftsgänge für Kauf und Pflichtsammlung umgesetzt oder möglich sein.

Der zu untersuchende Markt wird auf Programme eingegrenzt, die in deutschen Bibliotheken eingesetzt werden. Diese Begrenzung ist damit zu begründen, dass die Pflichtablieferung in den Ländern Europas und weltweit rechtlich und inhaltlich unterschiedlich ausgestaltet ist. Das kann sich in pflichtberechtigten Bibliotheken in sehr verschiedenen Geschäftsgängen spiegeln, die von Erwerbungsunterstützungsprogrammen entsprechend abgebildet werden und mit größerer Wahrscheinlichkeit in der Anpassung für die BSB-Geschäftsgänge weiteren Aufwand verursachen.

Recherchiert wurde in Fachbibliografien und Fachzeitschriften. Außerdem wurde eine Anfrage an die bibliothekarische Mailingliste Inetbib geschickt<sup>40</sup>. In der Vorauswahl schied eine auf Basis des Literaturverwaltungsprogramms Citavi programmierte Anwendung der Universitätsbibliothek Bamberg aus. Sie importiert Daten aus frei verfügbaren Quellen, bereitet sie auf und stellt sie für die Medienerwerbung der Fachreferent\*innen zur Verfügung. Diese Anwendung bietet jedoch keine Möglichkeit einer Selektion für die Pflichterwerbung.<sup>41</sup>

Im Ergebnis konnte nur ein Programm ermittelt werden, das die in der Anforderungsanalyse genannten Kriterien in einem Maße erfüllt, das eine Anpassung denkbar macht. Es wird im Folgenden näher untersucht.

---

<sup>40</sup> S. <http://inetbib.de/listenarchiv/msg66491.html> (zuletzt gesehen 6.12.2019). In Inetbib sind einige tausend Adressat\*innen vorwiegend aus dem deutschsprachigen Bibliothekswesen abonniert.

<sup>41</sup> Vgl. [Pierer, 2012].

## 5.2 eMAS, Württembergische Landesbibliothek

Vorbemerkung: eMAS konnte im Rahmen dieser Arbeit nicht live getestet werden. Die Darstellung des Systems folgt den beiden Artikeln, in denen seine Entwicklung und Funktionsweise beschrieben sind.<sup>42</sup> Ergänzend sind Informationen aus einer E-Mail vom 1.7.2019 und Telefongesprächen am 6.11.2019 und 25.11.2019 eingegangen. Die Informationen stammen jeweils aus dem Projektteam der WLB, das eMAS konzipiert und seine Entwicklung begleitet hat.

eMAS (elektronisches Medienauswertesystem; früher: Anawülas<sup>43</sup>) ist ein Erwerbungsunterstützungssystem, das die Württembergische Landesbibliothek (WLB) von 2015 bis 2017 im Rahmen eines Projekts für die eigene Medienerwerbung entwickelt hat.<sup>44</sup> Es handelt sich dabei um eine Single-Page-Webanwendung. Auf Serverseite wird ein HTTPS-gesicherter Apache-Webserver unter Linux/Ubuntu mit PHP 7 und MySQL eingesetzt. Auf Clientseite wird JavaScript/AJAX im Zusammenspiel mit HTML und CSS verwendet. Zur Kommunikation zwischen den Clients untereinander und mit dem Server werden AJAX und Websockets eingesetzt. Die Datenimporte der DNB-Titel erfolgen zeitgesteuert via crontab.

eMAS kann Daten aus verschiedenen Quellen und Formaten verarbeiten. Derzeit (Stand: 12.2019) werden die nationalbibliografischen Reihen A, B, O und H (Hochschulschriften) sowie Daten englischsprachiger Titel des Bibliothekslieferanten Dietmar Dreier, Duisburg, verarbeitet. Die Reihen werden wöchentlich bzw. monatlich im MARC21-Binärformat vom Bibliotheksservice-Zentrum Baden-Württemberg bezogen. Die Daten von Dietmar Dreier werden im Format MARC21 bezogen, intern aufbereitet und angereichert, in JSON umgewandelt und in die Datenbank importiert. Danach werden die Daten gegen den Bestand der WLB abgeglichen. Zum Abgleich werden interne Erwerbungsinformationen aus dem an der WLB eingesetzten System aDIS/BMS (Fa. aStec) benutzt. Dabei handelt es sich um Exemplar- und Titeldaten neu angelegter Medien mit weiteren Angaben (Erwerbungsstatus u. a.). Diese werden per PostgreSQL aus aDIS exportiert und in eine MySQL-Datenbank eingespielt. Durch Verwendung dieser internen Daten können nicht nur bereits vorhandene Titel,

---

<sup>42</sup> [Oberfell/Thalhofer, 2018], [Oberfell/Thalhofer, 2019].

<sup>43</sup> S. [WLB-Forum 2/2019], S. 45.

<sup>44</sup> Seit 2019 wird eMAS auch in der Badischen Landesbibliothek Karlsruhe eingesetzt (s. [WLB-Forum 2/2019], S. 45).

sondern z. B. auch erst bestellte Titel berücksichtigt werden. Der Bestandsabgleich erfolgt über ein mehrstufiges Verfahren anhand von Standardnummern (ISBN, ISSN u. a.), Erscheinungsjahr, Titel, Autor und anderen Angaben.

Danach werden die Titeldaten den zuständigen Bearbeitenden und Teams zugeordnet. Wie in der BSB gibt es in der WLB die Geschäftsgänge der Käuferwerbung über Referent\*innen und die Pflichterwerbung. Außerdem ist hier noch ein Geschäftsgang für die Landesbibliografie integriert, der an der BSB separat läuft.<sup>45</sup>

Daten der Reihe O werden derzeit nur in der Landesbibliographie zur Verfügung gestellt, nicht für die anderen Geschäftsgänge. Ein Einsatz in der Pflichterwerbung ist erst in Vorbereitung. Für die Käuferwerbung sind zunächst die Daten anderer Lieferanten eingeplant.

Die Zuordnung der Datensätze zu Fachreferent\*innen erfolgt über DDC-Sachgruppen. Für die Zuordnung zu Pflicht durchlaufen die Datensätze ein bis zu sechsstufiges Prüfverfahren, bei dem ISBN-Stämme baden-württembergischer Verlage, Verlagsorte, Verlagsnamen und in den Datensätzen vorhandene Länderkennzeichen Anwendung finden. Um Titel und Schlagwörter mit Baden-Württemberg-Bezug der Landesbibliografie zuzuordnen, werden sie mit einem in eMAS hinterlegten Index abgeglichen.

Allen Mitarbeitenden stehen die ihnen zugeordneten Titel auf einer Benutzeroberfläche zur Verfügung. Die Bearbeitung erfolgt über eine Menüleiste mit Funktionen, die von der Rolle der Mitarbeitenden abhängen. Nach Abschluss der Bearbeitung werden die Titelsätze der nächsten Station im Geschäftsgang zugeordnet. Ein Austausch von Titeln zwischen Geschäftsgängen wie auch zwischen Käuferreferent\*innen ist möglich.

Es gibt eine Oberfläche für Administrationszwecke; dort können Benutzer und Rollen verwaltet und verschiedene in eMAS hinterlegte Konkordanzen gepflegt werden.

Ein formales Lizenzmodell für eMAS besteht nicht. Die WLB ist bereit, eMAS kostenfrei an andere Bibliotheken weiterzugeben. Die WLB übernimmt ggf. im Rahmen einer Vereinbarung die kostenpflichtige Einrichtung des Systems für die

---

<sup>45</sup> Auf den ebenfalls integrierten Spezial-Geschäftsgang für die Internationale Hölderlin-Bibliographie soll hier nicht näher eingegangen werden.

lizenzierende Bibliothek. Eine Weiterentwicklung durch die WLB erfolgt bedarfsabhängig, kann aber auch durch die lizenzierende Bibliothek selbst geleistet werden.

Beim Abgleich mit der Anforderungsanalyse<sup>46</sup> ergeben sich folgende größere Anpassungen, die vor einem Einsatz von eMAS an der BSB vorzunehmen wären:

- Die Verbundzentrale des Bibliotheksverbands Bayern bietet den bayerischen Bibliotheken die nationalbibliografischen Reihen nicht als Binärdateien an, sondern im Format ASEQ (Aleph Sequential)<sup>47</sup>. Daher müsste der Datenimport angepasst werden.
- Bei der Bestandsprüfung zieht die WLB die benötigten Daten zum Abgleich aus ihrem lokalen Bibliothekssystem aDIS. Die BSB setzt ein anderes Bibliothekssystem ein (SISIS Sunrise); jedoch können die Bestandsinformationen für eMAS analog aus anderen Systemen gefüllt werden.

Problematischer ist, dass der Bestandsabgleich ab der bayerischen Pflichtgesetznovelle titelbezogen gegen den Bestand unterschiedlicher Bibliotheken ausgeführt werden muss (REQ-9). Diese Differenzierung müsste noch implementiert werden. Auch wäre ein automatisierter Datenabzug aus insgesamt neun lokalen Bibliothekssystemen nötig. Das würde einen erheblichen zusätzlichen Anpassungsaufwand bedeuten (gg. NF-REQ-6).

Für den Bestandsabgleich müsste die Exportroutine aus dem lokalen Bibliothekssystem vollkommen neu programmiert werden.

- Für die Zuordnung von Titeln zur Erwerbungsart Pflicht sowie zu unterschiedlichen zuständigen Bibliotheken (REQ-8) ist ab der Pflichtgesetznovelle eine Kombination zusätzlicher Kriterien zu den jetzt in eMAS verwendeten erforderlich. eMAS unterstützt in der Pflichterwerbung die Aufteilung auf zwei Zuständigkeitsbereiche (z. B. Regierungsbezirke); für die bayerische Situation würde die Aufteilung auf mehrere benötigt (REQ-6, NF-REQ-3). Die Berücksichtigung dieser Anforderungen würde ebenfalls eine Programmanpassung erfordern.

---

<sup>46</sup> S. Kap. 4.

<sup>47</sup> Ein Zeichenkettenformat, das zeilenbasiert serialisiert und vom Bibliothekssystem Ex Libris Aleph für den Import und Export von Datensätzen verwendet wird (s. <https://format.gbv.de/aseq>; zuletzt gesehen 6.12.2019).

– Das an der BSB nicht benötigte Modul Landesbibliografie kann im Betrieb de facto deaktiviert werden, indem der hinterlegte Index nicht gefüllt wird. Das Modul zu entfernen, wäre aufwändig.

Im Ergebnis ist festzuhalten, dass es sich bei eMAS um ein sehr differenziertes System handelt, das sehr gut an die Bedürfnisse einer wissenschaftlichen Bibliothek mit Kauf- und Pflichterwerbung angepasst ist. Zudem verfügt es über einige sehr attraktive Funktionen, z. B. einen Bestandsabgleich, der nicht nur ISBN und ISSN als Kriterien einsetzt. Dennoch hat die Analyse von eMAS gezeigt, dass einige Veränderungen nötig wären, um das Programm ausreichend an die Anforderungen der BSB anzupassen.

Um eine differenzierte Entscheidung für oder gegen eine Anpassung von eMAS treffen zu können, müsste zunächst dessen *reusability* analysiert werden, also die Möglichkeit, vorhandenen Code, Entwicklungsprozesse, Dokumente u. a. für einen Einsatz an der BSB anzupassen bzw. wiederzuverwenden. Außerdem müssten die Aufwände für die Anpassung einerseits und die Programmierung einer neuen Anwendung andererseits ermittelt und verglichen werden.

*Software reuse* ist eine Forschungsrichtung im Software Engineering mit eigener Methodik, die erst erarbeitet werden müsste. Zur praktischen Beurteilung der Aufwände ist darüber hinaus einige Erfahrung vonnöten. Nicht zuletzt bedeutet eine solche Analyse auch Zeitaufwand. All das ist im Rahmen dieser Arbeit nicht zu leisten. Es ist unwahrscheinlich, dass nach dieser Analyse in der zur Verfügung stehenden Zeit das Ziel der vorliegenden Arbeit noch zu erreichen wäre, nämlich durch Um- oder Neuprogrammierung ein für die Erwerbung der BSB einsetzbares Modul oder einen Prototyp zu erstellen.

Um dieses Ziel sicher erreichen zu können, wird daher entschieden, eine eigene Anwendung zu konzipieren und einen Prototyp mit begrenztem Umfang zu programmieren. Dieser Prototyp kann genau an die Bedürfnisse der BSB angepasst und künftig schrittweise ausgebaut werden.

### 5.3 Fazit

In dem definierten Markt konnte lediglich eine Softwarelösung ermittelt werden, die nach den Kriterien der Anforderungsanalyse als grundsätzlich geeignet zu betrachten ist, jedoch trotzdem einige Änderungen erfordert. Da eine detaillierte Einschätzung der Anpassungsaufwände nicht geleistet werden konnte, fiel die Entscheidung zugunsten der Konzeption einer eigenen Anwendung.

## 6. Konzept und Prototyp

### 6.1 Konzept

Aus der Anforderungsanalyse ergeben sich verschiedene Abläufe. Sie werden wie folgt zu größeren Abläufen zusammengefügt:

1. Datensätze der Reihe O werden importiert, gefiltert und in eine Datenbank gespeichert.
2. Datensätze werden vor einer Erwerbungsaktion<sup>48</sup> auf ihren Bestand an der zuständigen Bibliothek hin überprüft.
3. Datensätze werden den Referent\*innen der Käuferwerbung auf einer GUI zur Verfügung gestellt.
4. Datensätze werden an die zuständigen Erwerbungssteams weitergeleitet.

Neben diesen Abläufen sind eine Benutzeroberfläche und eine Datenbank zu implementieren.

#### Zu 1. Import

Als Datenquelle für die Reihe O wird die OAI-Schnittstelle der DNB verwendet. Über diese offene Schnittstelle stellt die DNB als Datenanbieterin strukturierte Metadaten zur Verfügung, die Diensteanbieter über *service requests* harvesten können. Das für

---

<sup>48</sup> Kaufentscheidung, Mahnung bei Pflichtverleger oder amtlichem Herausgeber.

die Kommunikation verwendete Protokoll heißt OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Die OAI-Schnittstelle eignet sich vor allem für den Abruf von Daten anhand von Zeiträumen, wie es die Erwerbungs geschäftsgänge der BSB erfordern.<sup>49</sup> Die Daten erhalten beim Abschluss der Verarbeitung durch die DNB einen Zeitstempel, anhand dessen sie abgefragt werden können.

Als Formate für den Datenbezug werden MARC 21-XML, DNB Casual und RDF angeboten.<sup>50</sup> MARC 21-XML ist das geeignete Format für das Erwerbungsunterstützungssystem, weil es die Inhalte der Datensätze vollständig abbildet. DNB Casual bietet hingegen nur eine Auswahl an Elementen (für Dublin Core) an. RDF ist für den Einsatz im Bereich Linked Data ausgerichtet.

Die SRU-Schnittstelle der DNB ist hingegen für die „gezielte Suche und Übernahme von Titel- und Normdaten“ bestimmt<sup>51</sup> und für die Abfrage der Gesamtdaten eines Zeitraums ungeeignet.

Eine weitere Möglichkeit des Datenbezugs wäre die Übernahme der Reihe O von der Verbundzentrale des Bibliotheksverbands Bayern im Format Aleph Sequential (ASEQ)<sup>52</sup>. ASEQ bietet jedoch keine Vorteile gegenüber XML. Durch die rein zeilenbasierte sequentielle Darstellung ohne menschenlesbare Elemente ist ASEQ im Gegenteil weniger komfortabel in der Programmierung des Parsers. Außerdem bezieht die Verbundzentrale die Daten der Reihe O nur monatlich, während der Bezug über die OAI-Schnittstelle flexibel möglich ist. Zudem ist mit der Verbundzentrale ein Dritter zwischen DNB und BSB geschaltet, was einen unnötigen Schritt im Geschäftsgang bedeuten würde.

Um Daten zu importieren, wird der gewünschte Zeitraum auf der GUI eingegeben und auf ‚Absenden‘ geklickt. Das Skript validiert die Eingaben client- und serverseitig. Dann prüft es tageweise gegen processedDays.txt, ob für diese Tage bereits Daten über die OAI-Schnittstelle abgefragt worden sind. Falls nein, ruft es die Importfunktion auf. Diese fragt die Daten bei der OAI-Schnittstelle der DNB ab. Die Datensätze

---

<sup>49</sup> S. <https://www.openarchives.org/pmh/>, [https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai\\_node.html](https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html) (zuletzt gesehen 6.12.2019).

<sup>50</sup> S. [https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai\\_node.html#doc58284bodyText9](https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html#doc58284bodyText9) (zuletzt gesehen 6.12.2019).

<sup>51</sup> [https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/SRU/sru\\_node.html#doc58294bodyText3](https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/SRU/sru_node.html#doc58294bodyText3) (zuletzt gesehen 6.12.2019).

<sup>52</sup> S. Fußnote 47.

werden in Teillieferungen zu je 50 Sätzen geschickt. Folgelieferungen können über einen mitgelieferten *resumption token* abgeholt werden.

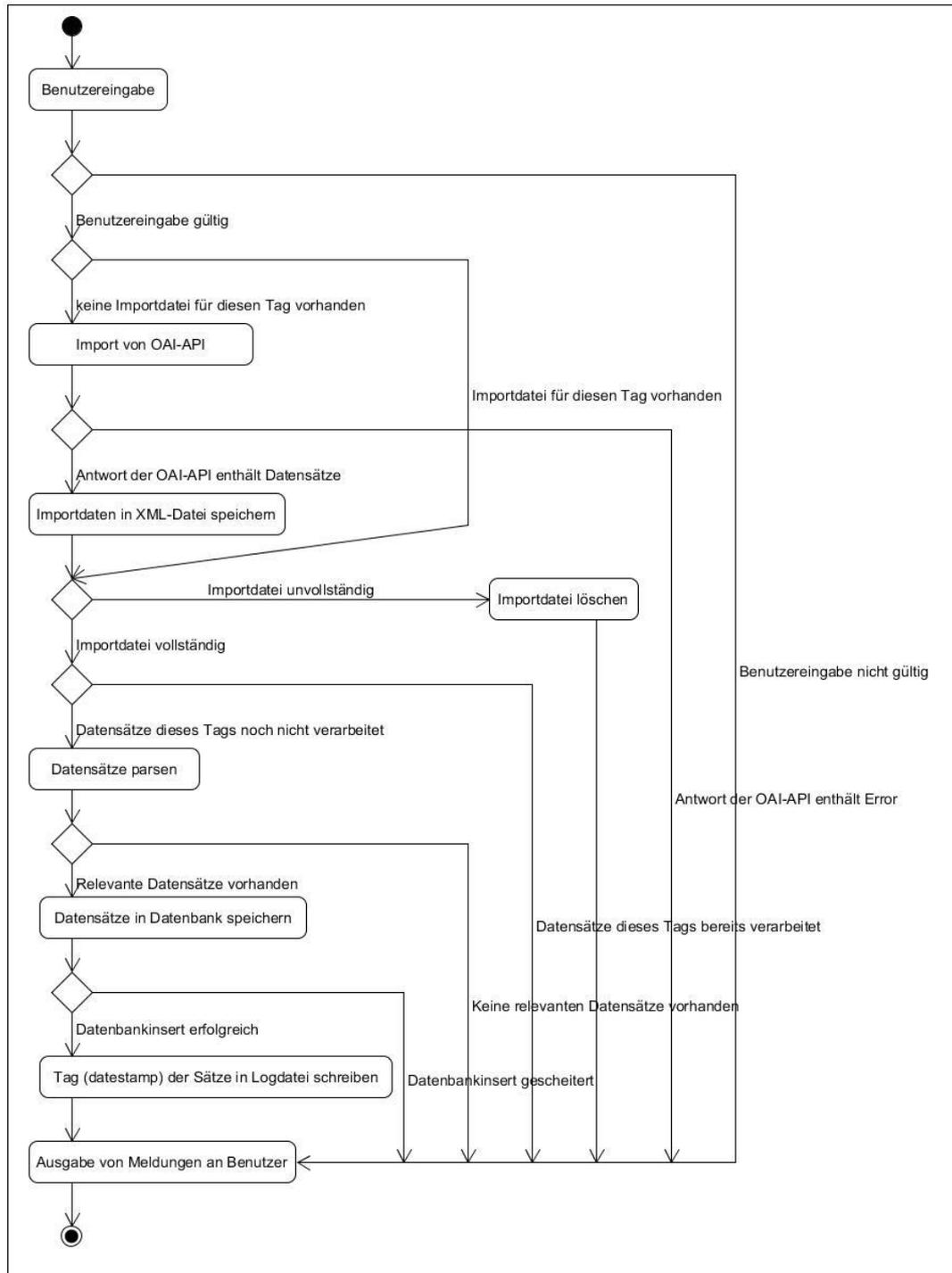


Abb. 3. Import (Aktivitätsdiagramm)

Die Teillieferungen werden in XML-Dateien gespeichert. Dabei werden alle Datensätze, deren Zeitstempel denselben Tag ausweist, in einer gemeinsamen Datei

zusammengefasst. Aus den Teillieferungen wird dabei ein wohlgeformtes XML-Dokument hergestellt. Eine Zusammenfassung nach Tagen ist vorteilhaft, da damit tageweise geprüft werden kann, ob Daten schon einmal abgeholt wurden (s. u., Prüfungen).

Nach dem Speichern wird geprüft, ob die XML-Dateien vollständig sind. Falls nicht, werden sie wieder gelöscht und der Durchlauf für diese Tage beendet. Falls ja, werden die XML-Dateien geparkt und aus den Datensätzen Datenobjekte erzeugt. Diese Objekte durchlaufen verschiedene Filter. Daten, deren Filterergebnis negativ ist, werden verworfen.

Positiv gefilterte Datensätze werden in die Datenbank gespeichert.<sup>53</sup> Damit wird ermöglicht, Datensätze zu unterschiedlichen Zeitpunkten weiterzuverarbeiten. So werden Pflicht- und amtliche Publikationen erst nach Ablauf von etwa zwei Monaten auf Bestand geprüft und ggf. bei den Produzenten gemahnt, wohingegen Kauftitel meist innerhalb weniger Tage nach Publikation durch die DNB bearbeitet werden.

(Aktivitätsdiagramm s. Abb. 3.)

### Zu 2. Bestandsprüfung

Die Bestandsprüfung kann für die drei Erwerbungsarten getrennt ausgeführt werden. Kaufreferent\*innen können sie zudem individuell nur für die ihnen – über Medientyp und DDC-Sachgruppen – zugeordneten Titel vornehmen.

Auf der GUI wird der gewünschte Zeitraum, ggf. noch Medientyp und DDC-Sachgruppen eingegeben, die Bestandsprüfung für eine der drei Erwerbungsarten ausgewählt und auf ‚Absenden‘ geklickt. Das Skript validiert die Eingaben und ruft bei positiver Prüfung die Bestandsprüfung auf.

Zunächst werden aus der Datenbank die Daten abgefragt. Selektionskriterien sind neben den oben genannten die folgenden:

- noch kein Bestand in der zuständigen Bibliothek
- noch nicht an das zuständige Erwerbsteam versandt.

---

<sup>53</sup> Datenbankschema s. Anhang A 7.

Dann wird für jede Standardnummer (ISBN oder ISSN) eine Anfrage an die SRU-Schnittstelle der B3Kat-Datenbank des Bibliotheksverbands Bayern (BVB) geschickt. Diese Datenbank enthält Titeldaten und Zeitschriften mit Besitznachweisen von Bibliotheken aus dem BVB und dem Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV).<sup>54</sup> Ist ein Titel mit dieser Standardnummer vorhanden, liefert die Schnittstelle eine XML-Seite zurück. Sie enthält in einem Datenfeld IDs derjenigen Bibliotheken, die den Titel mit dieser Standardnummer im Bestand haben. Ist die gesuchte Bibliotheks-ID dabei, wird für den Datensatz der Bestand in die Datenbank eingetragen.<sup>55</sup>

(Aktivitätsdiagramm s. Abb. 4.)

### Zu 3. Kaufauswahl

Referent\*innen geben auf der GUI Zeitraum, Medientyp und Sachgruppen der Titel ein, die sie für die Erwerbungsentscheidung anzeigen lassen wollen, und klicken auf ‚Titel zum Kauf auswählen‘. Das Skript validiert die Benutzereingaben und selektiert dann die angeforderten Datensätze aus der Datenbank. Zusätzlich zu den genannten, individuell eingegebenen Kriterien werden folgende Kriterien angewandt:

- Erwerbungsart ist Kauf
- Titel hat keinen Bestand in der BSB (da für die Bibliothek i. d. R. keine Mehrfachexemplare erworben werden, ist im gegenteiligen Fall eine Anzeige des Titels nicht mehr nötig)
- parallel in Print und elektronischer Form erscheinende Titel nur, wenn sie im bod-Verfahren publiziert werden (alle anderen parallel erscheinenden Titel werden auch in den Reihen A, B oder M angezeigt)
- das Datenbankfeld processingNotes enthält als Textanfang nicht ‚NOT‘ (s. u.)
- noch nicht an das zuständige Erwerbungssteam versandt

---

<sup>54</sup> S. <https://www.bib-bvb.de/web/b3kat/z39.50> (zuletzt gesehen 12.12.2019).

<sup>55</sup> Bei diesem Vorgehen wird genau genommen nicht geprüft, ob ein Titel an einer Bibliothek tatsächlich vorhanden ist. Abgefragt wird das sog. Besitzkennzeichen. Dieses wird auch dann vergeben, wenn eine Bibliothek einen Titel bestellt hat, er bestellt werden sollte, aber vergriffen gemeldet ist usw. In jedem Fall ist ein Besitzkennzeichen ein Nachweis, dass die Bibliothek bereits Kenntnis von diesem Titel hat und damit ihren Erwerbungsinteressen und -verpflichtungen nachkommen kann. Das ist in diesem Zusammenhang ausreichend.

## 6. Konzept und Prototyp

Die selektierten Titel werden auf der GUI zur Auswahl angezeigt. Die genaue Darstellung ist in Absprache mit den Referent\*innen zu gestalten.

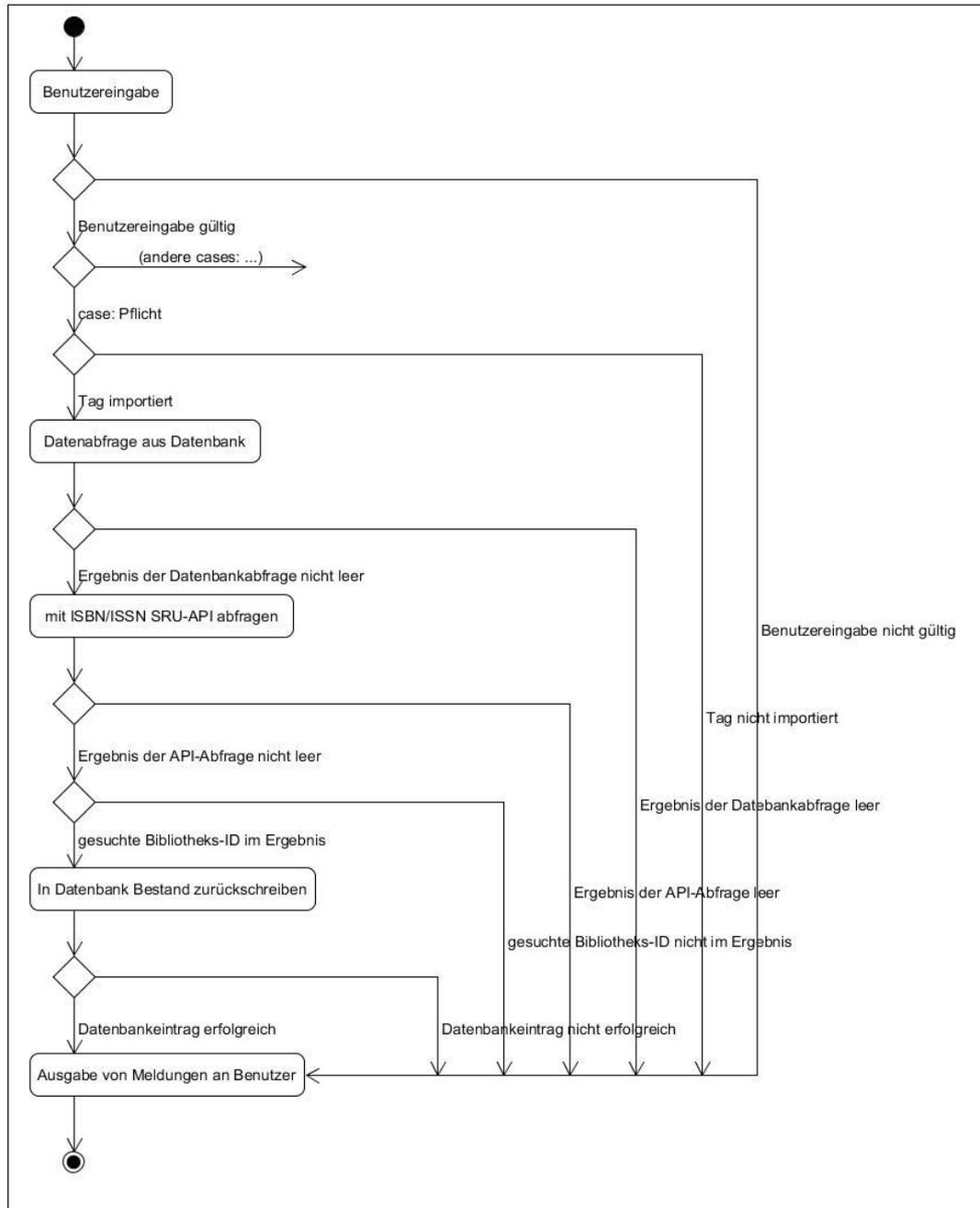


Abb. 4. Bestandsprüfung (Aktivitätsdiagramm)

Unter jedem Datensatz wird ein Eingabefeld angezeigt. Ist im Datenbankfeld `processingNotes` Text gespeichert, wird dieser im Eingabefeld als bearbeitbar angezeigt. In diesem Feld vermerken die Referent\*innen ihre Entscheidung. Wird ein Titel zur Erwerbung ausgewählt, ist als Textanfang ein standardisierter Text einzugeben (z. B. ‚Team Inland 1‘). Mithilfe dieses Textanfangs wird der Versand an die Kaufteams gesteuert. Wird als Textanfang ‚NOT‘ eingegeben, wird ein Satz endgültig abgelehnt. Beim Fortsetzen der Titelauswahl nach einer Unterbrechung mit Zwischenspeichern und erneutem Abruf der Sätze aus der Datenbank werden Sätze mit dieser Kennzeichnung nicht wieder angezeigt.

Klicken die Referent\*innen auf ‚Speichern‘, wird der Inhalt des Eingabefelds in die Datenbank gespeichert. Der Auswahlvorgang kann auf diese Weise auch unterbrochen und zu einem späteren Zeitpunkt fortgesetzt werden.

Um einen Datensatz an Kolleg\*innen weiterzuleiten, wird als Textanfang im Eingabefeld ebenfalls ein standardisierter Text eingegeben (z. B. ‚Balz‘ für ‚Nina Balz‘). Dieser Text wird beim Speichern aus dem Feldinhalt ausgelesen und in dem Attribut `firstDdc` gespeichert. In `firstDdc` wird standardmäßig die erste der DDC-Sachgruppen gespeichert; sie dient der Zuordnung von Titeln zu Referent\*innen. Mit diesem Verfahren ist es auch möglich, nach dem Namens Kürzel Bemerkungen für die Kolleg\*innen im Eingabefeld einzugeben.

Damit solche weitergeleiteten Titel ebenfalls angezeigt werden, geben die Referent\*innen im DDC-Sachgruppen-Feld zusätzlich zu den DDC-Sachgruppen ihr standardisiertes Namens Kürzel ein. Das Skript liest dann alle Datensätze aus, die im Feld `firstDdc` das Namens Kürzel enthalten. Im Unterschied zu den oben genannten Selektionskriterien wird dabei der angegebene Zeitraum nicht berücksichtigt, damit immer alle weitergeleiteten Datensätze, auch von länger zurückliegenden Tagen, ausgelesen werden.

(Aktivitätsdiagramm s. Abb. 5.)

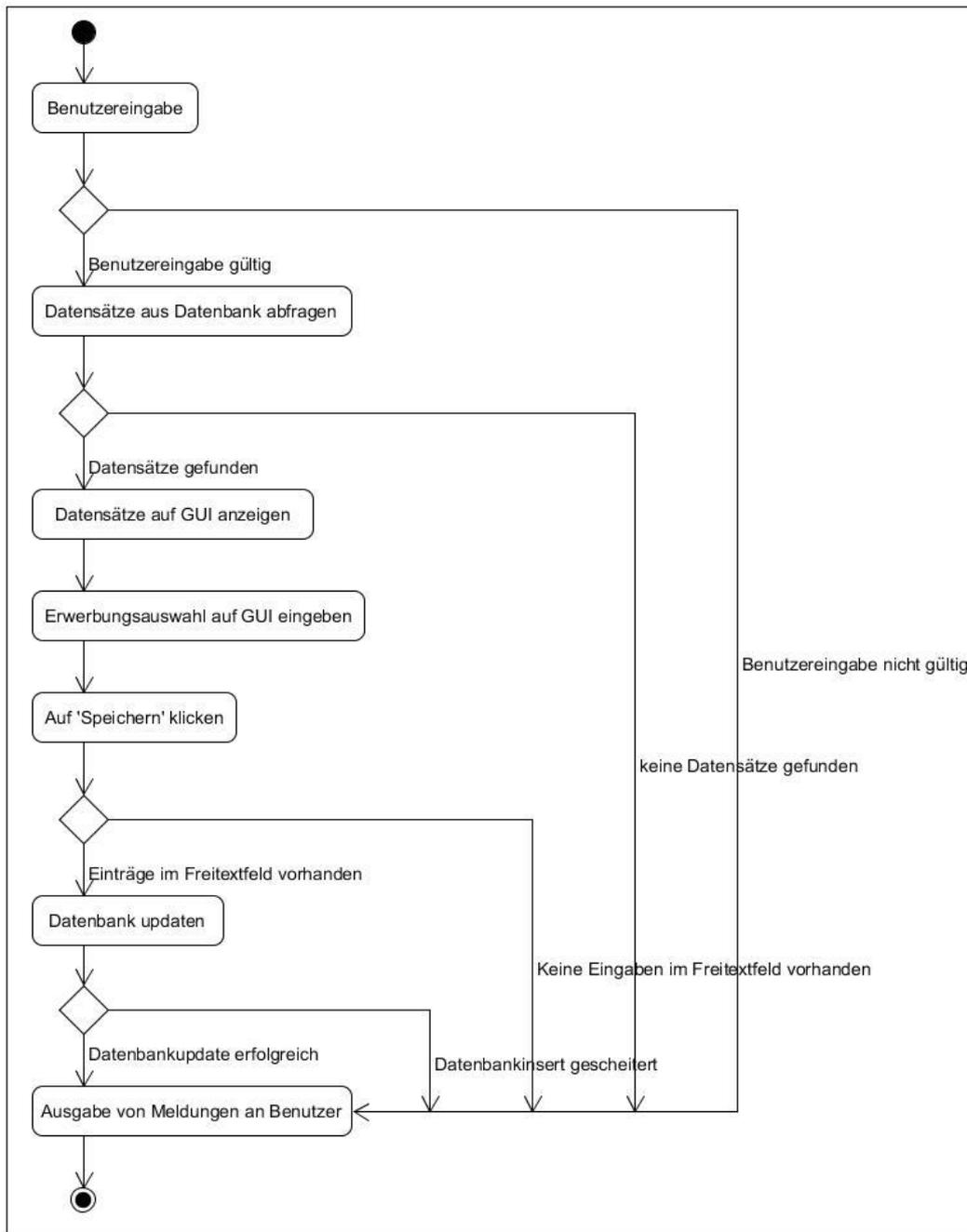


Abb. 5. Kauf (Aktivitätsdiagramm)

### Zu 4. Datenexport

Der Export der Daten für die Teams, die die weitere Bearbeitung übernehmen, erfolgt per E-Mail-Versand aus dem Skript heraus. Grund dafür ist, dass in der Medienbearbeitung der BSB bereits recht viele Programme und Arbeitsdateien eingesetzt werden. Durch anstehende Veränderungen in den Geschäftsgängen werden weitere Programme und Workflows hinzukommen. Die Teams sollen daher nicht mit einem zusätzlichen Programm arbeiten müssen, das lediglich die Reihe O-Workflows abdeckt. Aus demselben Grund werden auch aus dem ABM-Tool Daten per E-Mail exportiert. Damit wird für das Reihe-O-Tool ein gleichartiger Geschäftsgang aufgebaut.

Der Export wird von dem Referenten, der die Auswahl getroffen hat, bzw. bei Pflicht und Amtlichen Publikationen zentral von einer Person angestoßen. Um Datensätze zu exportieren, wird der gewünschte Zeitraum auf der Startseite eingegeben, für Kauftitel außerdem der Medientyp und die Sachgruppen, der Radiobutton ‚Pflichttitel ans Team schicken‘ (oder ‚Kauftitel...‘, ‚Amtliche Titel...‘) ausgewählt und auf ‚Absenden‘ geklickt.

Das Skript validiert zunächst die Benutzereingaben. Dann prüft es tageweise gegen processedDays.txt, ob bereits für alle gewünschten Tage Daten über die OAI-Schnittstelle abgefragt worden sind, und gegen sentToLegalDepositTeam.txt (bzw. eine entsprechende Datei für Amtliche Publikationen), ob sie bereits an das zuständige Team verschickt worden sind. – Für Kauftitel wird keine Prüfung gegen eine Versendet-Datei ausgeführt, weil Titel wiederholt und zu verschiedenen Zeitpunkten ausgewählt und zum Kauf an die Teams geschickt werden können.

Wurde ein Tag bereits abgefragt, eventuell importierte Daten jedoch noch nicht verschickt, wird eine Anfrage an die Datenbank geschickt und die Daten selektiert. Bei einem leeren Ergebnis bricht das Skript den Prozess mit einer Meldung ab.

Falls Daten vorhanden sind, werden sie in eine TXT-Datei geschrieben. Pflichttitel werden zusätzlich nach Regierungsbezirken sortiert und in unterschiedliche TXT-Dateien gespeichert, da die Bearbeitung nach Regierungsbezirken organisiert ist. Der Dateiname besteht aus einem Kürzel für die Erwerbungsart, bei Pflichttiteln danach einem Kürzel für den Regierungsbezirk sowie bei allen aus dem Datum der enthaltenen Datensätze (z. B. ‚PV\_obb\_2019-05-21, 2019-05-22.txt‘ für Pflichttitel aus dem Regierungsbezirk Oberbayern vom 21. und 22.5.2019).

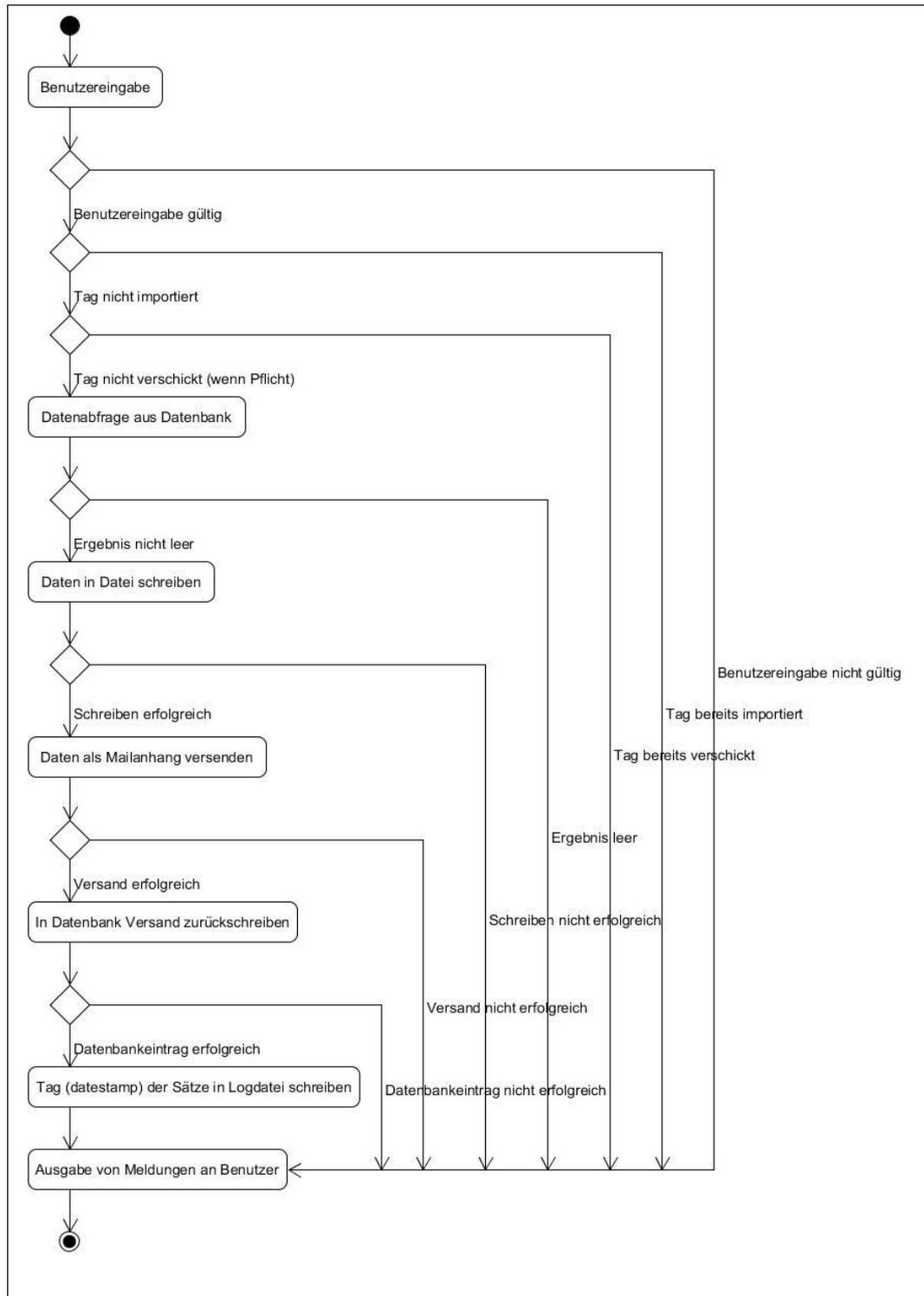


Abb. 6. Datenexport (Aktivitätsdiagramm)

Diese Datei wird als Anhang einer E-Mail an die Sammel-Mailadresse des zuständigen Teams versandt. Die Zuordnung zu Mailadressen erfolgt über die Erwerbungsart,

bei Kaufiteln zusätzlich über standardisierte Einträge in processingNotes. Im Betreff der E-Mail wird auch der Dateiname angegeben; damit können Sachbearbeiter\*innen die für sie bestimmten E-Mails schnell heraussuchen.

Ist der Versand erfolgreich, wird in der Datenbank für jeden versendeten Datensatz wasSentToTeam auf 1 gesetzt, und in die Logdatei sentToLegalDepositTeam.txt werden die bearbeiteten Tage eingetragen. Abschließend wird eine Meldung auf der GUI ausgegeben.

(Aktivitätsdiagramm s. Abb. 6.)

### Prüfungen

Im Programmablauf finden verschiedene Prüfungen statt, um zu verhindern, dass Prozesse unnötig wiederholt werden. So wird beim Importprozess geprüft, ob die Datensätze eines vom Benutzer angefragten Tags bereits erfolgreich durch das Skript bearbeitet wurden. (,Erfolgreich bearbeitet‘ ist nicht gleichbedeutend mit ,Datensätze dieses Tags wurden in die Datenbank gespeichert‘. Denn auch ein Import mit anschließender Filterung, die keine relevanten Datensätze ermittelt, ist eine erfolgreiche Bearbeitung.) Dazu wird geprüft, ob es einen Eintrag für diesen Tag in der Logdatei processedDays.txt gibt.<sup>56</sup>

Falls diese Prüfung negativ ist, wird geprüft, ob es für den Tag bereits eine Importdatei gibt. Falls nicht, wird der Import über die OAI-Schnittstelle gestartet. Falls doch, wird geprüft, ob die Importdatei vollständig ist. Eine unvollständige Datei wird gelöscht, und das Skript terminiert mit einer Fehlermeldung. Nach erneuter Benutzereingabe kann der Import erneut ausgeführt werden.

### Benutzeroberfläche

Sie dient dazu, mithilfe von Benutzereingaben die Prozesse Import, Bestandsprüfung und Datenversand zu starten und Erwerbungsentscheidungen für Kaufitel zu vermerken.

---

<sup>56</sup> Eine Prüfung gegen die Datenbank ist nicht sinnvoll möglich, da nicht für jeden abgerufenen Tag auch Datensätze in der Datenbank gespeichert werden. Zudem wäre das nicht so performant.

Tage, für die bereits erfolgreich eine Datenanfrage an die OAI-Schnittstelle geschickt worden ist, werden in einem Dropdownmenü in absteigender Reihenfolge angezeigt.<sup>57</sup>

Auf der Administrationsoberfläche können Importdateien gelöscht sowie Meldungstexte bearbeitet werden. Bei einer vollständigen Programmierung des Skripts sollen von hier aus auch Konkordanzen gepflegt werden können.

Die Administrationsseiten sind nicht gesondert durch ein Rechtemanagement o. ä. geschützt, um allen Beteiligten die Möglichkeit zu geben, Daten zu pflegen. Da das Skript im BSB-Intranet liegen wird und die URL nur den Anwendenden bekannt sein sollte, ist kein Schutz gegen böswillige Beschädigung vorzusehen. Sollte sich im Echtbetrieb herausstellen, dass ein Schutz doch benötigt wird, kann er nachträglich eingerichtet werden.

### Sicherheit

Wie geschildert, sind keine besonderen Maßnahmen gegen böswilligen Umgang mit dem Skript zu treffen. Als Sicherungsmaßnahmen wird das Skript im BSB-Intranet liegen, Benutzereingaben werden client- und serverseitig validiert, und die Kommunikation mit der Datenbank ist gegen SQL-Injektionen u. ä. zu sichern.

### **6.2 Softwareauswahl**

Wie in Kap. 6.1 gezeigt, besteht die zu konzipierende Anwendung aus den Elementen Datenimport und -export, Datenverarbeitung, Datenhaltung und grafische Benutzeroberfläche. Sie ist gut als Webanwendung umzusetzen, bei der die Datenverarbeitung auf einem Webserver stattfindet. Im Gegensatz zu Desktopanwendungen müssen Webanwendungen nicht auf jedem Rechner installiert werden, Änderungen können zentral durchgeführt werden, und da lediglich ein Webbrowser zur Benutzung benötigt wird, ist auch die Plattformunabhängigkeit vergleichsweise hoch. Damit werden Teile von NF-REQ-1 und NF-REQ-2 abgedeckt (Portabilität, einfache Wartung).

---

<sup>57</sup> Für Bestandsprüfung und Datenversand werden diese Informationen nicht angezeigt, weil beide Prozesse von mehreren Benutzer\*innen für unterschiedliche Teilmengen der Daten aufgerufen werden und im Kaufbereich auch wiederholt werden können.

Weitere Kriterien, die für die Softwareauswahl aus diesen nicht-funktionalen Anforderungen abgeleitet werden können, sind, dass die zu verwendende Software weit verbreitet und etabliert ist und nur wenige Abhängigkeiten enthält.

Wenn man betrachtet, welche Programmiersprachen für Webanwendungen eingesetzt werden, stehen auf den ersten Plätzen – je nach Untersuchungsfokus und Erhebungsmethode<sup>58</sup> – vor allem PHP, Java und Javascript. Die drei Sprachen haben eine Reihe gemeinsamer Merkmale und Qualitäten, auch wenn ihre Herkunft und Haupteinsatzgebiete differieren.

PHP wurde ausschließlich für die Webserverprogrammierung entwickelt und wird von den meisten Webservern automatisch unterstützt. Es wird von verschiedenen Betriebssystemen unterstützt und ist im World Wide Web weit verbreitet. Viele große Websysteme (z. B. Wordpress), aber auch die Mehrheit privater Webseiten sind in PHP implementiert.<sup>59</sup>

Java entstand 1991 und ermöglicht mit seinem Ansatz ‚write once, run everywhere‘ und der Kompilierung auf Bytecode, plattformunabhängige Programme zu schreiben. Vor allem große Unternehmen weltweit nutzen Java für ihre Geschäftsanwendungen im Web.<sup>60</sup> Durch den Einsatz für Android-Smartphones hat die Verbreitung von Java noch einmal deutlich zugenommen.

Javascript wurde ursprünglich für die clientseitige Programmierung zur Erweiterung von HTML und CSS entwickelt, kann aber inzwischen auch für die serverseitige Programmierung verwendet werden. Durch die Entwicklung der AJAX-Technologie wurde die Nutzung von Javascript neu belebt.<sup>61</sup>

---

<sup>58</sup> Nach der Verwendung auf Webseiten ohne weitere Einschränkung steht PHP weit vorne (s. [https://w3techs.com/technologies/overview/programming\\_language](https://w3techs.com/technologies/overview/programming_language), 6.12.2019: 78,9% PHP. Zur verwendeten Erhebungsmethode s. <https://w3techs.com/technologies>).

Bei weltweit beliebten Plattformen wie WordPress und Facebook sieht das Bild diverser aus (s. [https://en.wikipedia.org/wiki/Programming\\_languages\\_used\\_in\\_most\\_popular\\_websites](https://en.wikipedia.org/wiki/Programming_languages_used_in_most_popular_websites)).

Alle zuletzt gesehen 6.12.2019.

<sup>59</sup> S. [Herold et al., 2017], S. 506-507; [TB Informatik, 2012], S. 218; [Theis, 2018], S. 18-20.

<sup>60</sup> Lt. IDC; zitiert auf <https://www.yuhiro.de/technologien-fuer-die-entwicklung-von-webanwendungen/> (zuletzt gesehen 6.12.2019).

<sup>61</sup> S. [TB Informatik, 2012], S. 397.

Alle drei Sprachen können objektorientiert eingesetzt und für die Datenbankanbindung verwendet werden. Für alle gibt es eine große Zahl von Bibliotheken sowie spezialisierte Frameworks für Webanwendungen.<sup>62</sup>

Für die geplante Anwendung wurde PHP ausgewählt, da für diese Programmiersprache in der BSB Kompetenzen bestehen und auch das ABM-Tool in PHP geschrieben ist. Dadurch können bei gemeinsamer Wartung beider Anwendungen oder ggf. bei der Integration des ABM-Tools in das hier konzipierte Projekt<sup>63</sup> Synergieeffekte entstehen. Im Sinne einfacher Wartbarkeit wurde auf den Einsatz eines PHP-Frameworks wie Laravel oder Zend verzichtet, um ein Projekt mit wenigen Abhängigkeiten zu erstellen.

Zur Auswahl des Datenbanksystems sind die Art der zu speichernden Daten und der auf sie benötigte Zugriff zu betrachten. Bei den Daten handelt es sich um bibliografische Datensätze, d. h. um sehr gleichförmig strukturierte Daten. Nur auf wenige Attribute wird ein direkter Zugriff benötigt.

Damit kommt sowohl ein relationales als auch ein dokumentenorientiertes Datenbankmanagementsystem (DBMS) infrage. Unter den dokumentenorientierten Systemen ist besonders an MongoDB zu denken: Wie alle *document stores* kann dieses System unterschiedlich strukturierte Dokumente speichern und ist spezialisiert auf Big Data, hohe Skalierbarkeit und hohe Performanz. MongoDB ermöglicht zudem teilweise SQL-ähnliche Abfragen und nähert sich in diesem Punkt relationalen DBMS an. Abfragesprachen sind z. B. Java und PHP.<sup>64</sup>

Die Stärken von MongoDB werden für das vorliegende Projekt allerdings nicht benötigt: Die zu speichernden Daten haben eine gleichmäßige Struktur, und die anfallenden Datenmengen werden auch bei vollständiger Implementierung des Projekts im Echteinsatz nicht die Größenordnung von Big Data erreichen. Da MongoDB mit PHP angesprochen werden muss statt mit dem für diesen Zweck deutlich etablierteren SQL, wird, um dem Request der einfachen Wartbarkeit zu genügen, ein relationales Datenbankmanagementsystem eingesetzt.

---

<sup>62</sup> S. [Herold et al., 2017], S. 150, S. 506-507; [TB Informatik, 2012], S. 218; [Theis, 2018], S. 18-20.

<sup>63</sup> S. Kap. 8.2.

<sup>64</sup> S. [Edlich et al., 2011], S. 131-134, 140.

Aufgrund der bewährten Kombination von PHP mit MySQL bzw. MariaDB wird das Projekt mit XAMPP entwickelt. Bei XAMPP handelt es sich um eine freie Software, die den Apache-Server, MariaDB mit phpMyAdmin und PHP enthält (sowie weitere, hier nicht benötigte Komponenten).<sup>65</sup> XAMPP wird für Entwicklungszwecke breit eingesetzt, nicht zuletzt wegen der schnellen und einfachen Installation. Der Apache-Server, ein Freewareprodukt, wird für schätzungsweise zwei Drittel aller weltweit eingesetzten Webserver verwendet. Er wird fortlaufend verbessert und ist sehr gut dokumentiert.<sup>66</sup>

Zur Erzeugung der Benutzeroberfläche wird eine Templating Engine eingesetzt. Sie ermöglicht es, HTML- und PHP-Code sehr weitgehend bis vollständig zu trennen; dadurch wird der Code übersichtlicher. Auch können Anzeige und Applikationslogik unabhängig voneinander gewartet und weiterentwickelt werden.<sup>67</sup> Konkret wird die freie Open-Source-Software Smarty genutzt. Sie verfügt über eine einfache und klare, tag-basierte Syntax sowie an PHP angelehnte Kontrollstrukturen und Vererbung. Für Smarty spricht auch die hohe Verarbeitungsgeschwindigkeit, die durch Hintergrund-Kompilierung der Templates (und ggf. Caching) erreicht wird. Zudem ist sie plattform- und frameworkunabhängig.<sup>68</sup> Auch wenn Smarty bereits 2002 veröffentlicht wurde, ist sie nach Blade und Twig immer noch eine der von PHP-Programmierern meistgenutzten Templating Engines.<sup>69</sup>

Für die clientseitige Programmierung (v. a. Validierung) wird Javascript eingesetzt.

### 6.3 Entwurfsmuster und Klassenkonzept

Wie in Kap. 6.1 beschrieben, besteht die Anwendung aus den Elementen grafische Benutzeroberfläche, Datenhaltung sowie Datenimport und -export und interne Datenverarbeitung. Der Code kann durch Trennung nach Zuständigkeiten für diese Elementen

---

<sup>65</sup> S. <https://www.apachefriends.org/de/index.html> (zuletzt gesehen 6.12.2019).

<sup>66</sup> S. [TB Informatik, 2012], S. 398.

<sup>67</sup> S. z. B. [Maurice, 2010], S. 415; [Schlossnagle, 2006], S. 135-146.

<sup>68</sup> S. [Maurice, 2010], S. 415; [Schlossnagle, 2006], S. 135-142; [smarty.net/about\\_smarty](http://smarty.net/about_smarty).

<sup>69</sup> Mit 18%. Zum Vergleich: auf Platz 1 Blade mit 33%. 34% der befragten PHP-Programmierer gaben an, kein Framework einzusetzen. Laut Umfrage des Softwareunternehmens JetBrains im Jahr 2019; s. [https://www.jetbrains.com/lp/devecosystem-2019/php/?gclid=EAIaIQobChMIyKjdn8ra5AIVmM13Ch1vHwFcEAAYASAAEgIRa\\_D\\_BwE&gclsrc=aw.ds](https://www.jetbrains.com/lp/devecosystem-2019/php/?gclid=EAIaIQobChMIyKjdn8ra5AIVmM13Ch1vHwFcEAAYASAAEgIRa_D_BwE&gclsrc=aw.ds) (zuletzt gesehen 6.12.2019). – Anmerkung zur Statistik: Die Summe der Nennungen ergibt über 100%, da offenbar Mehrfachnennungen möglich waren.

te besser strukturiert werden. Dafür wird als Entwurfsmuster Model View Controller (MVC) eingesetzt. MVC trennt Präsentation (View), Steuerung (Controller) sowie Datenhaltung und -verarbeitung (Model) voneinander. Dadurch können diese Komponenten unabhängig voneinander weiterentwickelt werden. Auch die Wiederverwendbarkeit von Code wird damit unterstützt und allgemein Komplexität beherrschbar gemacht.<sup>70</sup> (Klassendiagramm s. Abb. 7.)

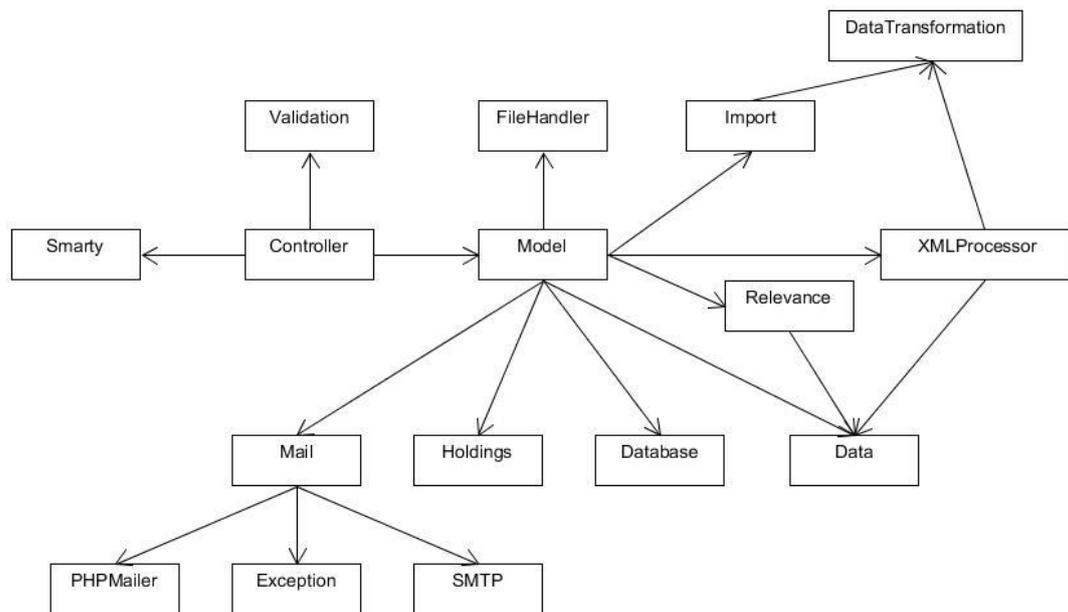


Abb. 7. Klassendiagramm

Die `index.php` dient dabei als Bootstrapdatei, die von der GUI eingehende Requests an den Controller weiterleitet. Der Controller vermittelt nur zwischen Model und View und hat selbst wenig Code. Er veranlasst darüber hinaus lediglich die Validierung der Benutzereingaben. Entsprechend verfügt der Controller neben dem Konstruktor nur über die `run`-Methode, in der über `cases` alle weiteren Funktionen angesteuert werden. (S. Abb. 8.)

---

<sup>70</sup> Vgl. [Geirhos, 2018], S. 519-541.

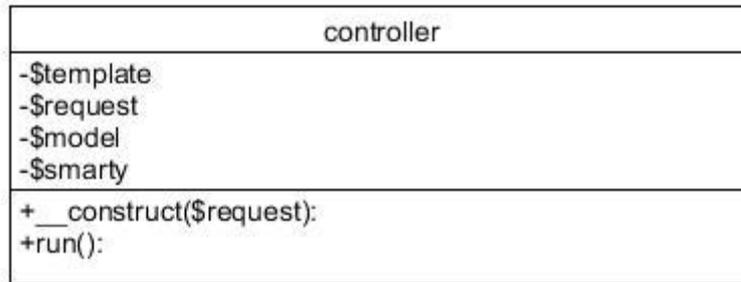


Abb. 8. Controller

Das Model nimmt die Anfragen vom Controller entgegen und arbeitet sie mithilfe der Business-Logic-Klassen und der Datenklasse (s. u.) ab. Das Model übernimmt v. a. das Zusammensetzen komplexer Funktionen aus einfacheren Funktionen der anderen Klassen. (S. Abb. 9.)

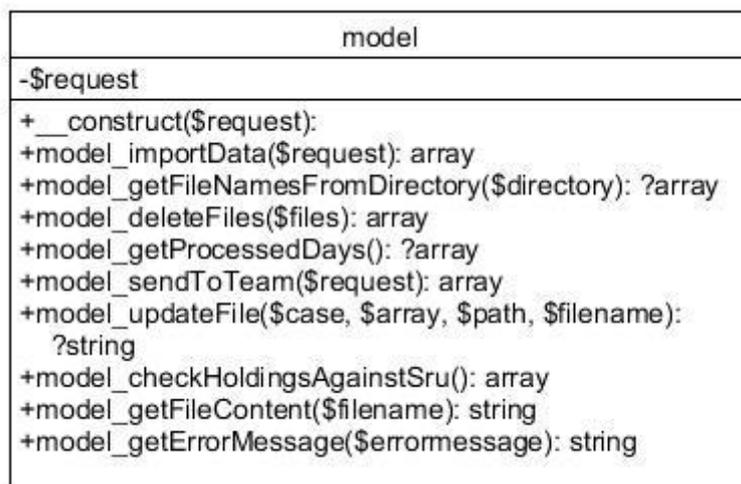


Abb. 9. Model

Die **Datenklasse Data** wird nur zur Datenspeicherung verwendet. Ihre Attribute korrespondieren den Spalten der Datenbanktabelle titles. Außer Settern und Gettern enthält Data zwei weitere Funktionen: `turnIntoAssociativeArrayForDatabaseInsert()` implementiert eine iterate-Funktion; sie gibt ein assoziatives Array zurück, das zum Insert in die Datenbank verwendet wird. `turnAssociativeArrayIntoDataObject($assocArray)` verwandelt umgekehrt ein assoziatives Array in ein Data-Objekt. (S. Abb. 10.)

Data
- \$titleId: string - \$title: string (...)
+ __construct(\$isInStockOfDepositLibrary = 'not tested', \$wasSentToTeam = false) + setTitleId(\$titleId): + setTitle(\$title): (...) + getTitleId(): string + getTitle(): string (...) + turnAssociativeArrayIntoDataObject(\$assocArray): Data + turnIntoAssociativeArray(): array

Abb. 10. Data (verkürzt dargestellt)

Die **Business Logic-Klassen** implementieren die Datenverarbeitung des Skripts. Sie verfügen selbst über keine Datenhaltung.

Die Klasse **Database** ist für die Kommunikation mit der Datenbank zuständig. Implementiert sind die Funktionen insert, update und select. In der weiteren Entwicklung ist noch eine delete-Funktion umzusetzen. Verwendet wird PDO (PHP Data Objects), eine objektorientierte Abstraktionsschicht für den Datenzugriff.<sup>71</sup> Prepared Statements mit Platzhaltern sichern den Code gegen SQL-Injections.<sup>72</sup> Inserts werden gesammelt mit einer Transaktion gespeichert. (S. Abb. 11.)

Database
+ update(\$tablename, \$changes, \$conditions): bool + insert(\$dataArray, \$tablename): array + select(\$fields, \$tablename, \$conditions): ?array

Abb. 11. Database

**DataTransformation** enthält Funktionen für verschiedene Manipulationen an Strings: adaptXML passt XML-Dateien so an, dass sie zusammen ein gültiges XML-Dokument bilden. getErrorText wandelt Systemmeldungen in für Anwendende

---

<sup>71</sup> S. <https://www.php.net/manual/en/intro.pdo.php> (zuletzt gesehen 6.12.2019).

<sup>72</sup> S. <https://www.php.net/manual/en/pdo.prepared-statements.php> (zuletzt gesehen 6.12.2019).

verständlichere Texte um. `turnCodeBitsToMediaType` ersetzt Kodierungen für Medientypen durch normalsprachliche Bezeichnungen.

`DataTransformation` soll um zusätzliche Funktionen erweitert werden, um z. B. weitere kodierte Angaben aus den bibliografischen Datensätzen (Sprachcodes u. a.) anhand von Konkordanzen in deutsche Bezeichnungen für die Ausgabe auf der GUI umzusetzen. (S. Abb. 12.)

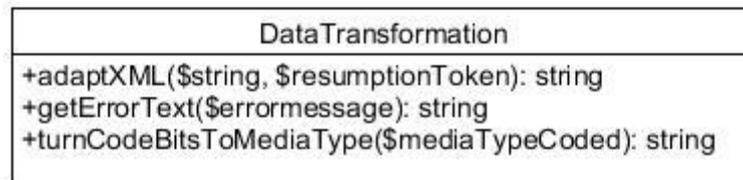


Abb. 12. *DataTransformation*

Die Klasse **FileHandler** übernimmt Aktionen im Zusammenhang mit Dateien. Es gibt Funktionen für das Schreiben in und Lesen aus Dateien, das Löschen von Dateien (intern von unvollständig importierten XML-Tageslieferungen sowie von der GUI aus von dort angezeigten Dateien) sowie zwei Funktionen, die Namen von Dateien ermitteln. (S. Abb. 13.)

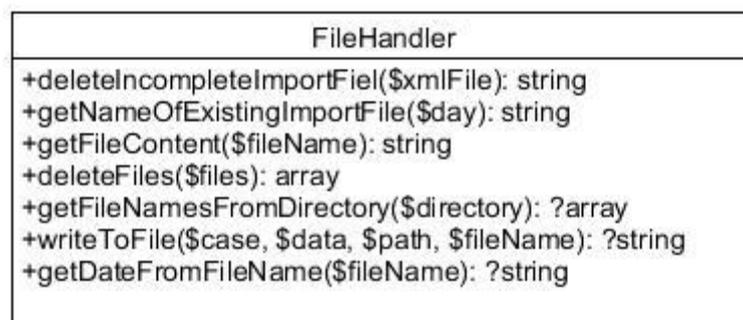


Abb. 13. *FileHandler*

Mit der Funktion `checkHoldingsAgainstSruApi` der Klasse **Holdings** können Datensätze anhand von einer Standardnummer (ISBN, ISSN) gegen den Bestand einer durch Bibliothekssigel identifizierten Bibliothek geprüft werden. Implementiert ist die Prüfung über eine SRU-Schnittstelle. (S. Abb. 14.)

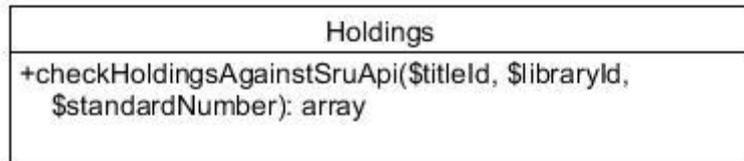


Abb. 14. Holdings

Die Klasse **Import** ist vor allem für den Datenimport verantwortlich. Implementiert ist mit `importFromOaiApi` die Abholung von Daten über eine OAI-Schnittstelle.

Mit der Funktion `wasProcessed` kann geprüft werden, ob Daten eines Tages bereits durch das Skript verarbeitet wurden. Der Parameter `$case` steuert, ob auf Import oder E-Mail-Versand für Pflichttitel geprüft werden soll.

Die Funktion `getProcessedDays` liest das Logfile `processedDays.txt` aus und gibt alle Tage zurück, für die schon Daten über OAI-API importiert wurden, für die Anzeige auf der GUI. (S. Abb. 15.)

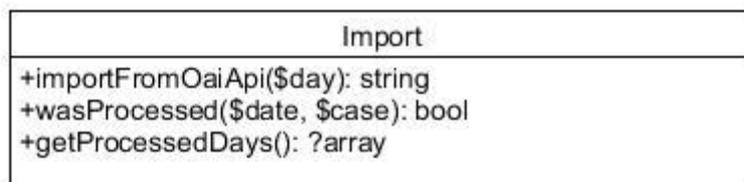


Abb. 15. Import

Die Klasse **Mail** setzt zum Versenden von E-Mails auf der Klasse `PHPMailer` von Brent R. Matzelle und Nachfolgern auf.<sup>73</sup> Die Funktion `send` unterscheidet je nach Inhalt des Parameters `$acquisitionType` verschiedene Fälle, nach denen Empfängeradresse, E-Mail-Text usw. gesetzt werden. Die mit `$fileName` angegebene Datei wird als Anhang versendet. (S. Abb. 16.)

---

<sup>73</sup> <https://github.com/PHPMailer/PHPMailer> (Download 11.10.2019).

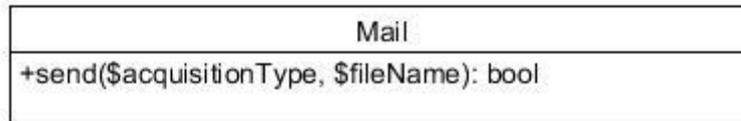


Abb. 16. Mail

Die Klasse **Relevance** implementiert mit der Funktion filter eine Prüfung für übergebene Data-Objekte, ob die Objekte für die weitere Verarbeitung herauszufiltern oder zu übergehen sind. filter enthält selbst einfache Kriterien und ruft für komplexere Prüfungen weitere, private Funktionen auf. Bei diesen Prüfungen können zusätzliche Werte für Attribute des Data-Objekts gesetzt werden. Bei der ersten positiv absolvierten Prüfung wird der weitere Durchlauf abgebrochen und je nachdem null oder das – ggf. angereicherte – Data-Objekt zurückgegeben. (S. Abb. 17.)

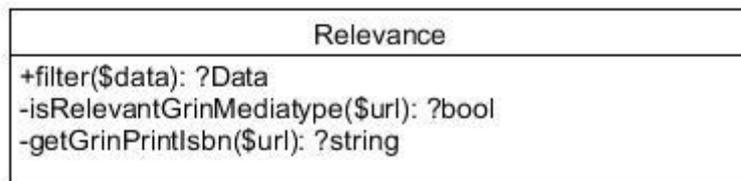


Abb. 17. Relevance

Die Klasse **Validation** ist für die Überprüfung der Eingaben zuständig, die Benutzer\*innen auf der GUI tätigen. Ein Request, der über index.php im Controller eingeht, wird von diesem an die Funktion validateUserInput weitergeleitet. Zunächst wird die übergebene Aktion auf Korrektheit überprüft. Abhängig von der gewählten Aktion werden ggf. weitere übergebene Parameter validiert. Die private Funktion preValidate ist zuständig für eine formale Vorprüfung aller Parameter, die privaten Funktionen isValidDate sowie isValidDateRange für die Prüfung von Zeitangaben.

Daneben überprüft die öffentliche Funktion checkForErrorMessage von der OAI-Schnittstelle zurückgelieferte Daten nach verschiedenen Patterns auf enthaltene Fehlermeldungen und gibt diese Meldungen oder einen leeren String zurück. (S. Abb. 18.)

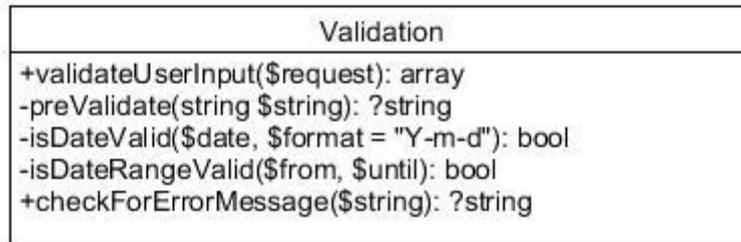


Abb. 18. Validation

Die Klasse **XMLProcessor** implementiert einen Expat-Parser. Der Parser liest ein Dokument abschnittsweise ein, durchläuft es zeilenweise und reagiert auf Events (öffnende und schließende Tags sowie Character Data). Er speichert Inhalte ausgewählter Subfields als Attribute eines Data-Objekts und gibt dieses Objekt zurück. (S. Abb. 19.)

Eine Tabelle mit der Konkordanz von Feldinhalten, XML-Tags und Subfields ist als Anhang A 5 enthalten.

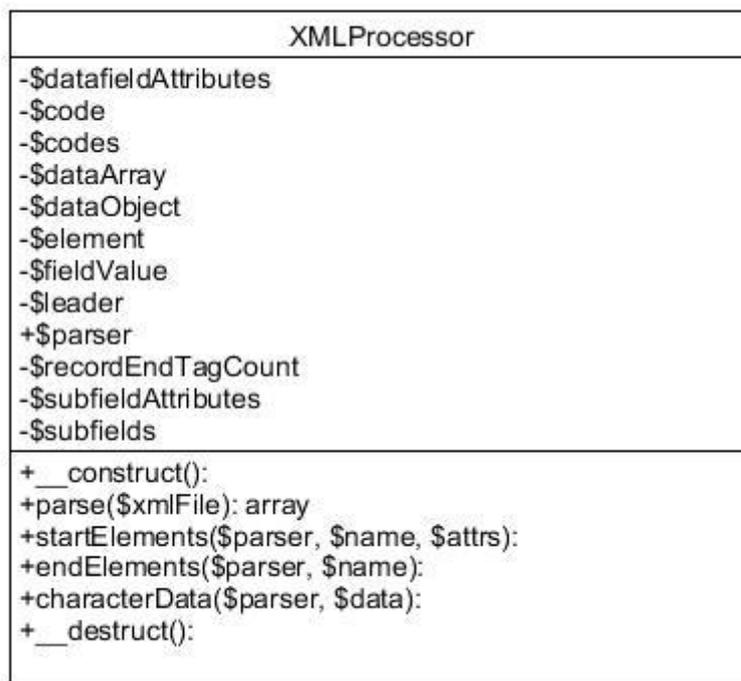


Abb. 19. XMLProcessor

### Ausgabe von Meldungen aus Funktionen; Exceptions

Um Benutzer über den erfolgreichen oder fehlerhaften Durchlauf des aufgerufenen Skriptteils zu informieren, werden Ausgaben auf der GUI erzeugt. Dazu geben Funktionen nach erfolgreichem Ablauf Texte (messages) zurück. Beim Auftreten von Exceptions oder anderer Fehler werden Fehlertexte (errormessages) zurückgegeben. Hierzu geben einige der vom Controller aufgerufenen Modelfunktionen ein Array aus einem messages-Array und einem errormessages-Array zurück. Im Controller werden die Inhalte beider Arrays getrennten Variablen im Template zugewiesen. Damit können beide auf der GUI getrennt und in unterschiedlicher Formatierung angezeigt werden. Fehlermeldungen werden über den Erfolgsmeldungen in roter Schrift ausgegeben, um mehr Aufmerksamkeit zu erzeugen.

Beim Werfen eigener Exceptions werden individuelle Texte und Fehlercodes eingesetzt (s. Anhang A 7). In der Klasse DataTransformation wurde mit der Funktion getErrorText die Möglichkeit geschaffen, Error- und Exception-Texte für die Ausgabe an Benutzer\*innen durch besser verständliche Texte zu ersetzen.

### **6.4 Weitere Dateien**

#### root-Verzeichnis

Autoloader.php lädt zur Laufzeit des Skripts benötigte Klassen dynamisch nach.

Doxyfile enthält die Einstellungen, um mithilfe von Doxygen eine PHP-Dokumentation zu erzeugen.

phpunit.phar enthält die Dateien für die Ausführung von PHPUnitTests.

#### Verzeichnis config

accessData.php enthält Zugangsdaten zur Reihe-O-Datenbank, URL und *Access Token* für die OAI-Schnittstelle sowie die URL für die SRU-Schnittstelle.

const.php enthält als Konstanten gespeicherte Werte für den Programmablauf, Texte für die Benutzeroberfläche und Meldungstexte, die den Benutzer\*innen ausgegeben werden.

emails.php speichert Daten und Texte für den E-Mail-Versand.

phpunit.xml ist die Konfigurationsdatei für PHPUnitTests.

### Verzeichnis html

Hier sind die Dateien der PHP-Dokumentation gespeichert.

### Verzeichnis importFiles

Das Verzeichnis speichert die über die OAI-Schnittstelle abgefragten Datensätze tageweise in XML-Dateien. Die Dateinamen folgen dem Schema ‚DNB\_OAI\_import\_YYYY-mm-dd.xml‘.

### Verzeichnis logFiles

processedDays.txt speichert zeilenweise die Tage, die bereits erfolgreich von der Anwendung verarbeitet wurden. Eingeschlossen sind Tage, für die es keine Daten oder keine positiv herausgefilterten Daten gibt. Ausgenommen sind Tage, bei deren Bearbeitung es zu Fehlermeldungen kam. Die Tage sind in der Form ‚YYYY-mm-dd‘ gespeichert.

sentToLegalDepositTeam.txt speichert in gleicher Weise die Tage, für die bereits zu mahnende Titel per Mail an das Pflichtteam geschickt wurden.

### Verzeichnis sentFiles

Hier werden die Dateien gespeichert, die als E-Mail-Anhänge versendet worden sind. Die Namen von Dateien für die Pflichtstelle folgen dem Schema ‚PV\_[Regierungsbezirk]\_YYYY-mm-dd.txt‘.

### Verzeichnis Smarty/templates

validation.js enthält die clientseitige Validierung mit JavaScript.

(Zum Verzeichnis concordances s. Kap. 6.6; zum Verzeichnis tests s. Kap. 9.4.)

## **6.5 Prototyp**

Zur Implementierung eines Prototyps wurden die Geschäftsgänge für die Pflichtablieferung ausgewählt. Dafür wurden die Abläufe ‚Datenimport‘, ‚Bestandsprüfung von Pflichttiteln‘ und ‚E-Mail-Versand von Pflichttiteln an das Pflichtteam‘ implementiert.

Wie in Kap. 2.2 dargestellt, sind Netzpublikationen bayerischer Verleger derzeit noch nicht an bayerische Pflichtbibliotheken ablieferungspflichtig. Insofern kann die Ablieferungskontrolle in der BSB für Pflichttitel im Wesentlichen mit den Reihen A, B und

M ausgeführt werden, die Printmedien nachweisen. Allerdings gibt es einen speziellen Aspekt der Pflichterwerbung, in dem auch bereits jetzt die Auswertung der Reihe O erforderlich ist. Medien, die parallel in gedruckter und elektronischer Form erscheinen und zudem im bod-Verfahren angeboten werden, werden von der DNB nur in der Reihe O vollständig nachgewiesen. Da diese Medien in ihrer gedruckten Form in Bayern der Pflichtablieferung unterliegen, muss zur Ablieferungskontrolle die Reihe O ausgewertet werden.

Derzeit gibt es nur einen bayerischen Verlag, der einen Teil seiner Produktion in dieser Weise anbietet.<sup>74</sup> Es handelt sich dabei um den GRIN-Verlag in München, dessen Titelangbot im sechsstelligen Bereich liegt<sup>75</sup>. Nicht alle Titel des Verlags fallen unter die Ablieferungspflicht, sondern nur Dissertationen, Habilitationsschriften, Fachbücher und Forschungsarbeiten. Allerdings werden von der DNB nur Dissertationen und Habilitationsschriften gekennzeichnet; ‚Fachbuch‘ und ‚Forschungsarbeit‘ sind Kategorien, die nur auf der Website des Verlags selbst zu recherchieren sind. Dort sind sie in den Metadaten der Einzelseiten hinterlegt, die es für jeden Titel gibt.

Daher wurde folgendes Vorgehen umgesetzt: Die Anwendung filtert alle Medien heraus, deren Medientyp E-Book, Zeitschriften oder Noten ist. Aus diesen werden diejenigen gefiltert, bei denen als Verlagsangabe ‚GRIN Verlag‘ hinterlegt ist, die als Book on Demand angeboten werden und in deren Datensatz ein Link auf eine GRIN-Verlagsseite enthalten ist, die Informationen zum jeweiligen Titel enthält. Über den Link wird die HTML-Seite abgefragt. Über Mustervergleich wird in den Metadaten geprüft, ob es sich um eine Dissertation, Habilitationsschrift, Forschungsarbeit oder ein Fachbuch handelt. Falls ja, wird in der Seite eine Print-ISBN gesucht.

Kommt es bei einer der Prüfungen zu einem negativen Ergebnis, wird die Filterung abgebrochen und der Datensatz verworfen. Bei erfolgreichem Durchlaufen aller Prüfungen wird die Print-ISBN im Datensatz gespeichert. Außerdem werden weitere Attribute des Satzes gefüllt (Erwerbungsart, Erwerbungssteam, zuständige Bibliothek, Regierungsbezirk).

---

<sup>74</sup> Alle anderen bayerischen Verlage, die Titel parallel in gedruckter und elektronischer Form anbieten, produzieren reguläre gedruckte Auflagen. Diese werden stets in den Reihen A, B oder M nachgewiesen.

<sup>75</sup> Mit Stand 6.12.2019 über 213.000 Titel (vgl. <https://www.grin.com>).

Da die Daten nicht über eine – *per definitionem* standardisierte und dokumentierte – Programmierschnittstelle abgefragt werden (können), ist dieses Vorgehen nicht vollständig robust. Zum einen kann sich die Struktur der URLs für die Einzeltitel-seiten ändern, wie es in den letzten Jahren bereits geschehen ist. Für den Fall erneuter Änderungen wurde die curl-Option FOLLOWLOCATION gesetzt. Sollte sich die URL-Struktur allerdings ändern, ohne dass der Verlag einen Redirect einrichtet, laufen die Anfragen dann jedoch trotzdem ins Leere.

Zum anderen sind die Bezeichnungen für die Medientypen nur bedingt standardisiert in den Verlagsseiten hinterlegt, denn sie sind abhängig von der Sprache des jeweiligen Titels. Wird eine Einzeltitelseite direkt aufgerufen und nicht durch Link innerhalb der GRIN-Site angesteuert, wird jeweils eine Seite in der Sprache des Titels ausgeliefert. Das betrifft z. T. auch die Metadaten. Damit mussten die Begriffe für die auszufilternden Medientypen in mehreren Sprachen in der Anwendung hinterlegt werden. Auch wird teilweise mehr als eine Bezeichnung je Sprache und Medientyp verwendet, z. B. in Englisch ‚Habilitation‘ und ‚Professorial Dissertation‘ für ‚Habilitationsschrift‘. Daher ist die Prüfung möglicherweise bereits jetzt unvollständig oder wird es später. Es ist somit in größeren Abständen zu kontrollieren, ob das Verfahren noch korrekt und vollständig funktioniert.

### **6.6 Hinweise zur weiteren Programmierung**

Codestellen, an denen bei Umsetzung des Konzepts Erweiterungen vorzunehmen sind, sind meist mit # HIER ERWEITERN # und z. T. weiteren Anmerkungen gekennzeichnet.

Um das Konzept vollständig zu implementieren, ist vor allem zweierlei zu tun: Zum einen sind Filterkriterien für Kaufmedien, Amtliche Publikationen und – nach der Novellierung des Pflichtstückegesetzes – weitere oder andere Kriterien umzusetzen (in der Klasse Relevance). Zum anderen sind Abläufe und Template für die intellektuelle Auswahl von Kauftiteln zu implementieren.

Für amtliche Publikationen ist ein Filter auf den amtlichen Herausgeber umzusetzen. Bei positiv herausgefilterten Publikationen ist ähnlich wie bei Pflichttiteln über ein Attribut im Datensatz festzulegen, in die Zuständigkeit welcher Sachbearbeiter\*innen

der jeweilige Titel fällt. Diese Zuständigkeiten können über Konkordanzen zwischen Herausgebern und Sachbearbeiter\*innen ermittelt werden (dafür ist ein Verzeichnis concordances vorgesehen).

Nach der Pflichtgesetznovelle wird die Sammlung bayerischer Netzpublikationen schrittweise aufgenommen werden. Begonnen wird voraussichtlich mit E-Books ausgewählter Verlage, deren Kreis allmählich ausgeweitet wird. Später können elektronische Zeitschriften hinzukommen usw. Die Filterung wird daher immer wieder zu verändern sein. Zu einem späteren Zeitpunkt, wenn alle in der Reihe O angezeigten Titel auch als Pflicht gesammelt werden, kann ein Filter für alle bayerischen Titel implementiert werden. Er sollte zuerst das Länderkennzeichen in den DNB-Sätzen abfragen und danach Listen mit bayerischen Verlagsorten, ISBN-Stämmen bayerischer Verlage sowie Verlagsnamen berücksichtigen.

Alle Titel, die weder Pflicht- noch amtliche Publikationen sind, durchlaufen die Filterung für Kauftitel. Die Relevanzkriterien sind noch von der bibliothekarischen Fachabteilung festzulegen.

Neben Relevanzfiltern können durch Filter Hinweise für die Referent\*innen erzeugt werden. Z. B. kann ein Hinweis auf eine Serie, die vollständig erworben werden soll, oder auf von der Bibliothek erworbene E-Book-Pakete eines Verlags generiert werden. Solche Hinweise werden in das Attribut `processingNotes` eingetragen und dann bei der Kaufauswahl im Freitextfeld angezeigt.

## 7. Kritische Bewertung

### 7.1 Laufzeit des Skripts

Der zeitaufwändigste Ablauf ist das Abholen und Parsen der XML-Dateien und der ggf. folgende Insert gefilterter Datensätze in die Datenbank, insbesondere im geplanten Regelbetrieb, bei dem Daten einer ganzen Woche auf einmal verarbeitet werden. Dabei wird in so gut wie allen Fällen die in der `php.ini` als Standard hinterlegte *maximum execution time* von 30 Sekunden deutlich überschritten. Die Verarbeitung einer

Wochenlieferung von durchschnittlicher Größe (ca. 15.200 Datensätze, 116 MB) dauerte im Test etwa 51 Minuten.<sup>76</sup>

Dieses Problem wurde dadurch gelöst, dass die Skriptlaufzeit im Skript selbst auf 0 gesetzt wird (über `ini_set()`). Der konkrete Wert ist in einer Konstante in `const.php` hinterlegt und kann über die Administrationsoberfläche verändert werden. Dieses Verfahren wurde einer Änderung von `max_execution_time` in `php.ini` vorgezogen, weil im Echtbetrieb auf derselben PHP-Installation voraussichtlich noch andere Skripte laufen, für die dieses Laufzeitverhalten nicht gelten soll.

Sollten später Sicherheitsbedenken gegen die Festsetzung der Laufzeit auf potentiell unendlich bestehen, ist zu empfehlen, die notwendige Sekundenzahl empirisch durch Abruf einer hohen Anzahl von Datensätzen zu ermitteln. Ein Wert von etwa 50.000 Datensätzen dürfte derzeit (Stand: Dezember 2019) als ausreichend anzunehmen sein. In den ersten drei Quartalen 2019 betrug der Spitzenwert für eine Woche knapp 38.000 Datensätze (Woche vom 25.2. bis 2.3.).<sup>77</sup>

Die Notwendigkeit dieser Manipulation macht deutlich, dass für den Datenimport der Einsatz von PHP keine optimale Wahl war. PHP ist, wie die als Standard voreingestellte *maximum execution time* verdeutlicht, an sich eine Programmiersprache für kürzer laufende Skripte. Eine Sprache wie Java, deren Prozesse unabhängig von der benötigten Zeit bis zum Ende durchlaufen und dann selbst terminieren, ist für solche langlaufenden Abläufe an sich besser geeignet.

Auch für die Person, die den Import über die GUI anstößt, ist diese lange Laufzeit nicht ideal, da erst danach Erfolg oder Misserfolg des Prozesses sichtbar wird. Daher sollte in der weiteren Entwicklung die Möglichkeit geschaffen werden, den Import automatisch über einen cronjob anzustoßen und Skriptrückmeldungen auch anders auszugeben. (S. Kap. 7.3, 7.4.)

---

<sup>76</sup> Rechnerkonfiguration: 8108 MB RAM, maximaler Prozessortakt 2,5 GHz, PHP-Version 7.2.3, für PHP zur Verfügung gestellter Arbeitsspeicher 128 MB (dynamisch per Skript: 256 MB).

<sup>77</sup> S. a. Anlage A 4.

## 7.2 Speicherverbrauch

Die Datenmengen, die pro Tag von der OAI-Schnittstelle der DNB abgerufen werden, sind recht umfangreich. Im Schnitt akkumulieren im betrachteten Zeitraum (1. bis 3. Quartal 2019; s. Anlage A 4) die Teillieferungen eines Tages zu einer XML-Datei mit etwa 18 MB. Einzelne Dateien werden aber auch deutlich größer als 30 MB. Die größte Datei ist fast 160 MB groß (27.2.2019), die größte bei Stichproben gefundene Datei 255 MB (1.12.2017). Eine durchschnittliche Wochenlieferung ist etwa 116 MB groß; die größte Wochenlieferung umfasst zusammen 255 MB (25.2.-2.3.2019).

Diese Dateigrößen belasten schnell den Festplattenspeicher. Derzeit können Dateien jedoch nur manuell – im Dateisystem oder von der Administrationsoberfläche aus – gelöscht werden. Als Verbesserung könnten Importdateien unmittelbar nach dem erfolgreichen Verarbeiten durch das Skript gelöscht werden. Alternativ könnten sie nach Ablauf einer festgelegten Zeitspanne beim nächsten dann folgenden Skriptaufruf oder mittels eines cronjobs gelöscht werden.

Für das Löschen von Datensätzen aus der Datenbank gibt es ebenfalls noch keine Funktion, da es nur sehr wenige erwerbungsrelevante Datensätze in den Lieferungen gibt. Später, wenn regelmäßig eine größere Anzahl von Datensätzen in die Datenbank importiert wird, sollte eine Funktion implementiert werden, die Datensätze nach ihrem Alter, das über das Attribut `datestamp` ermittelt werden kann, automatisch aus der Datenbank löscht. Auch dieses Löschen könnte beim Skriptaufruf oder mittels eines cronjobs angestoßen werden. (S. Kap. 7.3.)

Das Löschen älterer Importdateien und Datensätze ist für den Programmablauf unerheblich. Zum einen werden im Echtbetrieb Datensätze erfolgreich bearbeiteter Zeiträume nicht wieder abgefragt, und zum anderen bleibt der Eintrag dieser Tage in den Logdateien wie `processedDays.txt` für weitere Checks erhalten.

## 7.3 Mögliche Automatisierungsschritte

Wie in Kap. 7.1 und 7.2 dargestellt, sollten einige Funktionen eingeführt oder bereits existierende Funktionen so verändert werden, dass sie automatisiert und regelmäßig per cronjob angestoßen werden können: der Datenimport für wöchentliche Abfragen sowie das Löschen älterer Importdateien, Datenbankeinträge und Dateianhänge

versandter E-Mails. Außerdem können die Bestandsprüfung für Pflicht- und amtliche Publikationen sowie der Mailversand von Titeln, die in dieser Prüfung als nicht vorhanden ermittelt wurden, automatisiert werden.

### **7.4 Ausgabe von Systemmeldungen**

Meldungen des Skripts für Benutzer\*innen werden derzeit nur auf der GUI ausgegeben. Gerade auch angesichts dessen, dass v. a. Importe lange laufen können, bevor eine Meldung ausgegeben wird, ist es sinnvoll, Meldungen für Importe, Bestandsprüfungen und Mailversand ergänzend in eine Logdatei einzutragen. Erfolgs- und Fehlermeldungen können auch getrennt in zwei verschiedene Logfiles geschrieben werden. Diese werden im logFiles-Verzeichnis abgelegt und können von der Administrationsoberfläche aus eingesehen werden. Denkbar ist auch, sie zusätzlich jeweils bei neuen Einträgen per E-Mail an eine\*n Betreuer\*in zu versenden.

### **7.5 Implementierung des Controllers**

Die Entscheidung, für alle Views nur einen gemeinsamen Controller einzusetzen, hat sich als zu optimistisch erwiesen. Bereits für die beiden umgesetzten Views (Aktionsauswahl, Administrationsoberfläche) mussten zwölf cases in der run()-Methode des Controllers implementiert werden, was diese unübersichtlich gemacht hat.

Es sollte je View ein eigener Controller eingesetzt und zur Steuerung ein übergeordneter Frontcontroller implementiert werden.

### **7.6 Datenhaltung**

Die Struktur der Datenhaltung in der Datenbank ist zu überprüfen. Zum einen könnte es performanter sein, einige Attribute, die als Selektionskriterien verwendet werden (datestamp, acquisitionType u. a.), in eigene Tabellen – jeweils mit zugeordneten titleIds – auszulagern, um den Zugriff auf die benötigten Datensätze zu beschleunigen. Entsprechende Tests müssen mit größeren Datenmengen ausgeführt werden.

Zum anderen ist es unnötig, Datenfelder, auf die nicht separat zugegriffen werden muss, in eigenen Spalten zu speichern (z. B. additionalInfo, costOfAccess, digital Preservation). Sie könnten auch als assoziative Arrays in einem Array zusammengefasst und serialisiert in einer gemeinsamen Spalte der Datenbank gespeichert werden.

### 7.7 Kodierung der Daten

Es ist nicht gelungen, aus den Importdateien einheitlich mit UTF-8-kodierte Daten zu extrahieren, obwohl die XML-Dateien der DNB bereits mit UTF-8 kodiert sind.<sup>78</sup> Der Expatparser erkennt laut Dokumentation die verwendete Kodierung automatisch und gibt, wenn nichts anderes angegeben ist, die eingelesenen Strings UTF-8-kodiert aus.<sup>79</sup> Abweichend davon zerteilt der implementierte Parser die Strings in ASCII- und UTF-8-kodierte Strings und gibt sie einzeln aus. Die Teilstrings müssen daher im XMLProcessor wieder zusammengesetzt werden. Der Versuch, nur die ASCII-(Teil-)Strings oder die zusammengesetzten Strings in gemischter Kodierung nach UTF-8 umzukodieren (über mb\_detect\_encoding(), utf8\_encode() ), schlug fehl. Zusätzlich wurde für den Import in die Datenbank UTF-8 als Kodierung gesetzt, auch das ergebnislos.

Allerdings wurden bislang keine Probleme bei der Zeichendarstellung beobachtet. Das ist nicht weiter erstaunlich, da ASCII ein Subset von UTF-8 ist und ASCII-kodierte Zeichen damit in gültiger Weise auch UTF-8-kodiert sind. Ein Arbeiten mit der derzeitigen Kodierungssituation ist daher möglich.

---

<sup>78</sup> <?xml version="1.0" encoding="UTF-8"?>.

<sup>79</sup>S. die Dokumentation zur Kodierung auf <https://www.php.net/manual/en/function.xml-parser-create.php> (zuletzt gesehen 6.12.2019).

## 8. Ausblick

### 8.1 Integration anderer nationalbibliografischer Reihen

Bereits jetzt kann das Skript auch Daten anderer nationalbibliografischer Reihen verarbeiten. Dazu muss lediglich in `accessData.php` die Konstante `SET` auf den gewünschten Wert gesetzt werden, z. B. auf `,dnb:vv:reiheM'`. Auch das Parsen der XML-Importdateien muss nicht grundsätzlich verändert werden, da sich die Feld- und Unterfeldstruktur von MARC 21-XML für unterschiedliche Reihen nicht unterscheidet.

Um jedoch mehrere Reihen parallel über das Skript verwalten zu können, müssen folgende Änderungen vorgenommen werden:

Auf der Benutzeroberfläche wird zusätzlich die Eingabe der abzufragenden Reihe gefordert. – Der Controller reicht dieses Attribut über das `$_POST`-Superglobal-Array weiter ans Model; hier muss nichts angepasst werden.

In folgenden Klassen müssen die Schnittstellen um einen Parameter `,Reihe'` erweitert und z. T. weitere Anpassungen vorgenommen werden:

#### Import

- Funktion `import`: Damit kann über den `SET`-Bestandteil der URL die richtige Reihe von der OAI-Schnittstelle abgefragt werden.
- Funktion `wasProcessed`: Dann kann das passende Logfile abgefragt werden.

#### FileHandler

- Funktion `writeToFile`: (falls die Ausgabe der Daten je nach Reihe differenziert werden soll, z. B. für Noten)

#### Mail

- Funktion `sendMail`: `Betreff`, `Body-Text` usw. sind anzupassen.

#### Relevance

- Funktion `filter`: Die Relevanzkriterien sind zu differenzieren.

Für jede Reihe müssen eigene `importFiles` wie auch `logFiles` in separaten Verzeichnissen abgelegt werden.

In der Datenbank ist für jede Reihe eine separate Tabelle anzulegen. Damit wird der Zugriff gegenüber der Speicherung aller Daten in einer einzigen Tabelle beschleunigt, da stets nur Daten einer Reihe auf einmal bearbeitet werden.

## **8.2 Anpassung an die Bedürfnisse anderer Bibliotheken**

Das Skript kann auf einfache Weise an die Bedürfnisse anderer Bibliotheken angepasst werden, da es durchgängig objektorientiert programmiert ist, Zugangsdaten usw. zentral in config-Dateien hält und keine kostenpflichtigen Dienste oder lokalspezifische Systeme einsetzt. Voraussetzung ist lediglich, dass die Bibliothek ihre Bestände in einem Online-Katalog nachweist, der über eine Schnittstelle abzufragen ist (eigener Katalog, Verbundkatalog o. ä.).

Wird der Bestandsabgleich nicht über eine SRU-Schnittstelle ausgeführt, ist eine neue Funktion in der Klasse Holdings zu implementieren.

Die Kriterien, nach denen Datensätze gefiltert werden, können individuell in der Funktion filter der Klasse Relevance hinterlegt werden.

Werden weitere Datenfelder aus den Datensätzen benötigt, ist der XMLProcessor anzupassen.

Für Bibliotheken ohne Pflichtrecht oder Sammlung amtlicher Publikationen werden die nicht benötigten Aktionen auf der index.php ausgeblendet.

## Anhang

### A 1 Glossar

Aktion	Funktion auf der GUI des Reihe-O-Tools, z. B. Datenimport
Amtliche Publikation	Publikation, die von einem amtlichen Herausgeber veröffentlicht worden ist
Amtlicher Herausgeber	Einrichtungen, Stiftungen (usw.) öffentlichen Rechts, die Publikationen herausgeben
Bibliothekssystem	Von Bibliotheken für interne Geschäftsabläufe wie Erwerbung, Katalogisierung und Ausleihe eingesetzte Software
Erwerbungsart	Art und Weise, auf die ein Medium in den Bestand einer Bibliothek kommt. Für diese Arbeit sind folgende Erwerbungsarten relevant: Kauf, Pflichtsammmlung (nach dem bayerischen Pflichtstückgesetz), Erwerbung amtlicher Publikationen.
ErwerbungsTeams	Teams, die Medien für den Bestand der Bibliothek erwerben, per Kauf, Lizenzierung, als Pflichtmedium u. a.
Kaufentscheidung	Vermerk von Kaufreferent*innen, dass ein Titel gekauft werden soll. Er kann folgende Textelemente enthalten: Fachnummer, Sammelschwerpunkt, ErwerbungsTeam, weitere frei formulierte Bemerkungen (z. B. ‚Benutzerwunsch‘, ‚eilt‘).
Kaufreferent*innen	Mitarbeitende der Bibliothek, die entscheiden, welche Medien für den Bestand der Bibliothek gekauft werden sollen
KaufTeam	ErwerbungsTeam, das Medien im Kauf erwirbt
Pflichtstelle	ErwerbungsTeam, das die Pflichtablieferung bayerischer Verlage kontrolliert, säumige Ablieferer mahnt und eingesandte Medien bearbeitet
Unberechneter Zugang	Sachgebiet, das alle Medien ab ca. Erscheinungsjahr 1950 bearbeitet, die ohne direkte Erwerbungskosten in die BSB kommen. Ein Team im Unberechneten Zugang ist die Pflichtstelle.

Zuständige Bibliothek	<p>Bei den Erwerbungsarten Kauf und amtliche Publikationen ist die Bayerische Staatsbibliothek zuständige Bibliothek.</p> <p>Bei der Erwerbungsart Pflicht ist bis zur Novellierung des Pflichtstückegesetzes stets auch die BSB zuständige Bibliothek.<sup>80</sup></p> <p>Ab der Novellierung wird abhängig vom Sitz eines Verlags eine Bibliothek im selben Regierungsbezirk zuständige Bibliothek sein. Ausnahme: Für Noten wird stets die Hochschule für Musik, München, zuständige Bibliothek sein.</p>
-----------------------	---

## A 2 Filterkriterien für Datensätze

Kriterium	Beispiel	Relevant für Erwerbungsart
Amtlicher Herausgeber	Landesamt für Datenschutz	Amtliche Publikationen
Bayerischer Erscheinungsort	Regensburg	Pflichterwerbung
Dateiformat	PDF, epub	Kauf, Pflichterwerbung, amtliche Publikationen
DDC-Sachgruppe	650	Kauf
Erscheinen eines Titels parallel in elektronischer und körperlicher Form	--	Kauf, Pflichterwerbung
Herstellung eines Titels im bod-Verfahren	--	Pflichterwerbung
ISBN-Stamm eines bayerischen Verlags	978-3-406-	Pflichterwerbung
Kostenpflichtigkeit	kostenpflichtig, lizenzpflichtig	Kauf
Medientypen	Zeitschrift	Kauf, Pflichterwerbung, amtliche Publikationen

<sup>80</sup> Vgl. [Pflichtstückesammlung, 1986], S. 530.

<b>Kriterium</b>	<b>Beispiel</b>	<b>Relevant für Erwerbungsart</b>
Name eines bayerischen Verlags	C.H.Beck	Pflichterwerbung

### **A 3 Zuordnungskriterien für Datensätze der Pflichterwerbung zu einer zuständigen Bibliothek**

<b>Kriterium</b>	<b>Beispiel</b>
Verlagsort	Amberg
Medientyp	Zeitschrift
Publikationsweise	Book-on-Demand
Parallele Erscheinungsweise in gedruckter und elektronischer Form	--

Die Zuordnungskriterien werden ab dem Zeitpunkt der Novellierung des bayerischen Pflichtstückegesetzes benötigt. Die Zuordnung von Datensätzen zu einer zuständigen Bibliothek erfolgt über eine Kombination der Kriterien.

### **A 4 Konkordanz Feldinhalt - XML-Tags – Attribute der Data-Klasse**

<b>Feldinhalt</b>	<b>XML-Tags</b>	<b>Attribut der Data-Klasse</b>
<i><b>Allgemein</b></i>		
Satz-ID der DNB	controlfield tag="001"	\$titleId
<i><b>Bibliografische Daten</b></i>		
Allgemeine Anmerkung – zusätzliche Angaben, z. B. Abstract –	datafield tag="500" ind1=" " ind2=" " subfield code="a"	\$additionalInfo

Beteiligte Person/en	<p>datafield tag="100" ind1="0" ind2=" " subfield code="a"</p> <p>datafield tag="100" ind1="1" ind2=" " subfield code="a"</p> <p>datafield tag="700" ind1="0" ind2=" " subfield code="a"</p> <p>datafield tag="700" ind1="1" ind2=" " subfield code="a"</p>	\$author
Übergeordnetes Gesamtwerk	<p>datafield tag="490" ind1="0" ind2=" " subfield code="a" (<b>Titel</b>), subfield code="v" (<b>Bandangabe</b>)</p>	\$collectiveTitle
Ausgabebezeichnung	<p>datafield tag="250" ind1=" " ind2=" " subfield code="a"</p>	\$edition
Dateiformat	<p>datafield tag="856" ind1="4" ind2=" " subfield code="q"</p>	\$format
Hochschulschriftenvermerk	<p>datafield tag="502" ind1=" " ind2=" " subfield code="a"</p> <p>datafield tag="502" ind1=" " ind2=" " subfield code="b", subfield code="c", subfield code="d"</p>	\$highSchoolPublication
Sprache	<p>datafield tag="041 " ind1=" " ind2=" " subfield code="a"</p>	\$language
Seitenzahl	<p>datafield tag="300" ind1=" " ind2=" " subfield code="a"</p>	\$numberOfPages
Beteiligte Organisation	<p>datafield tag="110" ind1="1" ind2=" "</p>	\$organisation

	subfield code="a", subfield code="b"  datafield tag="710" ind1="1" ind2=" " subfield code="a", subfield code="b"	
Publikationsdatum, Ausgabebezeichnung, Seitenzahl (bei Periodika)	datafield tag="264" ind1=" " ind2="1" subfield code="c" <b>(bei Monografien)</b>  datafield tag="773" ind1="1" ind2=" " subfield code="g": "day: ", "month: ", "year: ", "volume: ", "number: ", "pages: "	\$publicationDate
Verlag	datafield tag="264" ind1=" " ind2="1" subfield code="b"	\$publisher
Verlagsort	datafield tag="264" ind1=" " ind2="1" subfield code="a"	\$publishingPlace
Standardnummer: ISBN, ISSN	datafield tag="020" ind1=" " ind2=" " subfield code="a" <b>(ISBN)</b> ,  datafield tag="022" ind1=" " ind2=" " subfield code="a" <b>(ISSN)</b>	\$standardNumber
Titel	datafield tag="245" ind1="0" ind2="0" subfield code="a" <b>(Titel)</b> , subfield code="b" <b>(Untertitel)</b>	\$title
<b><i>Sacherschließung</i></b>		
DDC-Sachgruppe/n der DNB	datafield tag="082" ind1="7" ind2="4" subfield code="a"	\$ddc

Freie Schlagwörter	datafield tag="653" ind1=" " ind2=" " subfield code="a"	\$freeSubjectHeadings
GND-Schlagwörter	(vorgesehen für künftige Auswertung) <sup>81</sup>	\$gndSubjectHeadings
<b>Zugangs- und Lizenzinformationen</b>		
Hinweise zu Zugangsbeschränkungen	datafield tag="506" ind1="0" ind2=" " subfield code="a"	\$accessInfo
Rechte/Zugriff auf das Original mit/ohne Kosten	datafield tag="856" ind1="4" ind2=" " subfield code="z"	\$costOfAccess
Info zur Langzeitarchi- vierung bei der DNB	datafield tag="856" ind1=" " ind2="0" subfield code="x"	\$digitalPreservation
Link zu diesem Datensatz bei der DNB	datafield tag="856" ind1=" " ind2="0" subfield code="u"	\$dnbUrl
Lizenzinformationen/ Copyright	datafield tag="540" ind1=" " ind2=" " subfield code="a"	\$licenceInfo
Persistent Identifier	datafield tag="856" ind1="4" ind2=" " subfield code="u" <b>(Persistent Identifier)</b> , subfield code="x" ( <b>Verlag o. ä.</b> )	\$persistentIdentifier
<b>Informationen für den Programmablauf</b>		
Erwerbungsart	--	\$acquisitionType
Bundesland	datafield tag="044" ind1=" " ind2=" " subfield code="c"	\$laenderkennzeichen

<sup>81</sup> Zur Zeit gibt es noch keine verbale Sacherschließung nach RSWK o. ä. in der Reihe O (s. <https://www.basiswissen-rda.de/neues-sacherschliessungskonzept-faz/>; zuletzt gesehen 6.12.2019).

Zeitstempel (Fertigstellungsdatum des Satzes bei der DNB)	header datestamp	\$dateStamp
Vorhanden in der Pflichtbibliothek?	--	\$isInStockOfDepositLibrary
Parallel in Print und online publiziert?	datafield tag="776" ind1="0" ind2="8" subfield="n": "Druck-Ausgabe"	\$isPublishedPrintAndOnline
Als Book-on-Demand publiziert?	datafield tag="925" ind1="p" ind2=" " subfield code="a": pd	\$isPublishedOnDemand
Bibliotheks-ID	--	\$libraryId
Medientyp	leader: Position 06-07, controlfield tag="008" Position 21, 24-27	\$mediaType
Erwerbungsentscheidung (bei Kauftiteln), Bearbeitungsvermerk (bei Pflichttiteln)	--	\$processingNote
Regierungsbezirk	--	\$regierungsbezirk
Team (bei Kauftiteln), Sachbearbeiter (bei Pflichttiteln)	--	\$responsibleTeamOrColleague
Ans Team geschickt?	--	\$wasSentToTeam

### A 5 Kodierung der Medientypen in Leader und Tag 008

Medientyp	Leader Pos. 06 + 07 <sup>82</sup>	Tag 008 Pos. 21 *	Tag 008 Pos. 24-27 *
Hörbuch	i* (2. Stelle egal)	(egal)	
Jahrbuch (Titel)	as	(egal)	yyyy

<sup>82</sup> Gezählt wird von Position 00 an.

<b>Medientyp</b>	<b>Leader Pos. 06 + 07<sup>82</sup></b>	<b>Tag 008 Pos. 21 *</b>	<b>Tag 008 Pos. 24-27 *</b>
Monografie	am	(egal)	--
Noten (monographisch)	cm	(egal)	
Notenserie (Titel; gezählte Serie)	cs	(egal)	
Notenserie (Titel; ungezählte Serie)	cm	(egal)	
Nummerierte Serie (Titel)	as	m	
Serienstück (einer nummerierten Serie)	ab	(egal)	
Serienstück (einer unnummerierten Serie)	am	(egal)	
Zeitschrift (Titel)	as	p	
Zeitschriftenheft	ab	(egal)	
Zeitung (Titel)	as	n	
Website	ai	(egal)	

## A 6 Exceptions

<b>Fehlercode</b>	<b>Kurzbeschreibung</b>	<b>Fehlertext (Konstante)</b>
1	Logfile kann nicht ausgelesen werden	PROBLEM_WITH_LOG_FILE
2	Von der OAI-API zurückgeliefertes Dokument enthält eine Fehlermeldung	ANSWER_OAI_API
3	Keine Datei übergeben	NO_FILE_GIVEN
4	Datei leer	FILE_EMPTY

<b>Fehlercode</b>	<b>Kurzbeschreibung</b>	<b>Fehlertext (Konstante)</b>
5	Datei kann nicht geöffnet werden	CAN_T_OPEN_FILE
6	In die Datei kann nicht geschrieben werden	CAN_T_WRITE_TO_FILE
7	Fehler beim Insert in die Datenbank	(verschiedene Fehlertexte)
8	Validierungsproblem	CHOOSE_CORRECT_ACTION, STARTDATE_FORMAT_WRONG, ENDDATE_FORMAT_WRONG, UNPLAUSIBLE_DATERANGE
10	Keine Internet- verbindung	NO_INTERNET_CONNECTION
11	Keine Datei zum Löschen ausgewählt	NO_FILE_GIVEN_TO_DELETE
12	Kein Datum übergeben	NO_DATE_GIVEN
13	Kein Eintrag ins Logfile	NO_ENTRY_INTO_LOGFILE
14	Datenbank-Update nicht möglich	DATABASE_UPDATE_NOT_POSSIBLE
15	Die Datei existiert nicht (mehr).	FILE_DOESN_T_EXIST
16	Das Verzeichnis existiert nicht.	NO_DIR
17	Keine Title-ID übergeben.	NO_TITLE_ID_GIVEN
18	Keine E-Mail-Adresse für das Pflichtteam hinterlegt	NO_EMAIL_RECIPIENT_PV
19	Fehlende Einstellung/en für den E-Mail-Versand	(verschiedene Fehlertexte)

## A 7 Datenbankschema: Tabelle titles

Spalte	Typ	Attribute	Null	Standard	Extra	Verweise	Kommentare	MIME
titleid	Int(20)		Nein					
title	varchar(500)		ja	NULL				
additionalInfo	blob		ja	NULL				
author	blob		ja	NULL				
collectiveTitle	blob		ja	NULL				
edition	varchar(50)		ja	NULL				
format	varchar(50)		ja	NULL				
highSchoolPublication	varchar(100)		ja	NULL				
language	blob		ja	NULL				
mediaType	varchar(50)		ja	NULL				
numberOfPages	varchar(20)		ja	NULL				
organisation	blob		ja	NULL				
publicationDate	varchar(20)		ja	NULL				
publisher	blob		ja	NULL				
publishingPlace	blob		ja	NULL				
standardNumber	blob		ja	NULL				
ddc	blob		ja	NULL				
freeSubjectHeadings	blob		ja	NULL				
gndSubjectHeadings	varchar(100)		ja	NULL				
accessInfo	varchar(100)		ja	NULL				
costOfAccess	varchar(20)		ja	NULL				
digitalPreservation	varchar(50)		ja	NULL				
dnbUrl	varchar(40)		ja	NULL				
licenceInfo	varchar(50)		ja	NULL				
persistentIdentifier	varchar(50)		ja	NULL				
laenderkennzeichen	varchar(10)		ja	NULL				
datestamp	varchar(10)		ja	NULL				
isInStockOfDepositLibrary	tinyInt(1)		ja	NULL				
isPublishedPrintAndOnline	tinyInt(1)		ja	NULL				
processingNote	varchar(50)		ja	NULL				
wasSentToTeam	tinyInt(1)		ja	NULL				
firstDdc	varchar(10)		ja	NULL				
acquisitionType	varchar(10)		ja	NULL				
acquisitionTeam	varchar(10)		ja	NULL				
libraryId	varchar(30)		ja	NULL				
regierungsbezirk	varchar(10)		ja	NULL				
isPublishedOnDemand	tinyInt(1)		ja	NULL				

**A 8 Datensätze und Dateigrößen**

(1.1.-30.9.2019)

Minimum: 53 Datensätze (0,31 MB)

Maximum: 23.889 Datensätze (151 MB)

Durchschnittlich 2.531 Datensätze pro Arbeitstag

Durchschnittlich 15.200 Datensätze pro Woche (Montag – Samstag)

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
01.01.2019	482	2.534	
02.01.2019	677		5,5
03.01.2019	631		4,2
04.01.2019	583		4,1
05.01.2019	161		1,2
06.01.2019			
07.01.2019	580	7.861	3,8
08.01.2019	1.117		7,7
09.01.2019	2.708		13,6
10.01.2019	1.053		7,7
11.01.2019	2.121		14,3
12.01.2019	282		2,0
13.01.2019			
14.01.2019	1.990	7.321	14,5
15.01.2019	1.613		15,4
16.01.2019	187		1,8
17.01.2019	1.923		12,9
18.01.2019	714		4,9
19.01.2019	894		7,1
20.01.2019			
21.01.2019	4.910	16.128	34,5
22.01.2019	6.175		50,4
23.01.2019	1.738		12,8
24.01.2019	919		7,6
25.01.2019	2.163		15,1
26.01.2019	223		1,6
27.01.2019			
28.01.2019	3.746	15.092	30,5
29.01.2019	3.393		24,8
30.01.2019	2.642		20,6

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
31.01.2019	2.136		14,6
01.02.2019	2.094		16,8
02.02.2019	1.081		8,8
03.02.2019			
04.02.2019	5.284	17.410	36,5
05.02.2019	2.115		16,9
06.02.2019	2.685		19,1
07.02.2019	1.816		12,2
08.02.2019	3.943		27,6
09.02.2019	1.567		13,0
10.02.2019			
11.02.2019	11.396	29.221	82,6
12.02.2019	3.983		27,5
13.02.2019	3.623		23,9
14.02.2019	3.836		25,9
15.02.2019	3.441		27,6
16.02.2019	2.942		22,3
17.02.2019			
18.02.2019	7.276	21.612	49,6
19.02.2019	3.445		24,1
20.02.2019	2.159		15,9
21.02.2019	4.877		35,7
22.02.2019	2.948		20,9
23.02.2019	907		7,3
24.02.2019			
25.02.2019	5.147	37.931	38,6
26.02.2019	2.520		18,3
27.02.2019	23.889		151,0
28.02.2019	2.786		20,6
01.03.2019	2.469		17,1
02.03.2019	1.120		9,1
03.03.2019			
04.03.2019	7.084	18.059	32,7
05.03.2019	3.696		27,1
06.03.2019	1.500		12,4
07.03.2019	1.112		4,3
08.03.2019	3.249		22,8
09.03.2019	1.418		11,6
10.03.2019			
11.03.2019	8.180	24.061	59,3
12.03.2019	4.842		32,9

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
13.03.2019	2.590		18,2
14.03.2019	2.403		16,3
15.03.2019	4.596		41,3
16.03.2019	1.450		12,2
17.03.2019			
18.03.2019	8.108	20.604	57,0
19.03.2019	3.539		27,0
20.03.2019	2.129		15,4
21.03.2019	2.928		21,2
22.03.2019	3.321		24,6
23.03.2019	579		4,4
24.03.2019			
25.03.2019	8.925	18.977	60,0
26.03.2019	1.817		13,2
27.03.2019	2.064		17,1
28.03.2019	3.189		22,5
29.03.2019	2.221		15,4
30.03.2019	761		6,0
31.03.2019			
01.04.2019	4.748	28.420	30,5
02.04.2019	2.978		22,6
03.04.2019	4.151		29,9
04.04.2019	3.729		27,6
05.04.2019	11.236		61,6
06.04.2019	1.578		13,2
07.04.2019			
08.04.2019	6.707	22.009	44,4
09.04.2019	4.187		26,0
10.04.2019	4.102		24,9
11.04.2019	2.681		18,1
12.04.2019	3.106		20,9
13.04.2019	1.226		9,8
14.04.2019			
15.04.2019	7.209	20.629	46,6
16.04.2019	3.412		23,4
17.04.2019	3.625		24,5
18.04.2019	2.845		18,9
19.04.2019	2.646		17,0
20.04.2019	892		7,2
21.04.2019			
22.04.2019	5.718	16.433	34,1

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
23.04.2019	2.701		17,6
24.04.2019	2.991		19,8
25.04.2019	1.431		12,8
26.04.2019	2.236		14,3
27.04.2019	1.356		11,5
28.04.2019			
29.04.2019	6.825	19.814	42,6
30.04.2019	2.462		16,0
01.05.2019	3.800		11,8
02.05.2019	3.120		2,4
03.05.2019	2.704		7,1
04.05.2019	903		7,3
05.05.2019			
06.05.2019	6.389	16.960	39,3
07.05.2019	2.464		16,0
08.05.2019	2.496		16,2
09.05.2019	2.382		14,5
10.05.2019	2.948		20,9
11.05.2019	281		2,3
12.05.2019			
13.05.2019	4.525	20.655	27,9
14.05.2019	3.063		21,1
15.05.2019	3.399		21,9
16.05.2019	3.245		21,7
17.05.2019	4.471		32,3
18.05.2019	1.952		15,7
19.05.2019			
20.05.2019	11.687	17.219	40,2
21.05.2019	1.845		11,5
22.05.2019	2.511		23,6
23.05.2019	808		5,4
24.05.2019	368		2,7
25.05.2019			
26.05.2019			
27.05.2019	1.837	22.537	11,4
28.05.2019	900		7,6
29.05.2019	9.644		61,1
30.05.2019	4.515		31,4
31.05.2019	4.743		30,5
01.06.2019	845		6,4
02.06.2019	53		0,3

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
03.06.2019	2.662	11.164	17,4
04.06.2019	2.035		13,5
05.06.2019	1.332		9,6
06.06.2019	1.781		11,6
07.06.2019	2.209		14,4
08.06.2019	1.145		6,9
09.06.2019			
10.06.2019	1.928	8.465	11,7
11.06.2019	799		6,2
12.06.2019	2.066		14,7
13.06.2019	1.882		14,2
14.06.2019	1.207		9,5
15.06.2019	583		4,3
16.06.2019			
17.06.2019	2.519	7.436	21,6
18.06.2019	1.210		9,3
19.06.2019	786		6,6
20.06.2019	1.642		12,3
21.06.2019	640		4,7
22.06.2019	639		4,7
23.06.2019			
24.06.2019	2.442	13.805	19,4
25.06.2019	1.152		9,3
26.06.2019	5.368		49,6
27.06.2019	1.833		13,9
28.06.2019	1.616		11,7
29.06.2019	1.394		11,0
30.06.2019			
01.07.2019	1.800	11.736	14,0
02.07.2019	1.415		12,0
03.07.2019	1.702		13,5
04.07.2019	3.932		26,7
05.07.2019	1.545		11,3
06.07.2019	1.342		10,5
07.07.2019			
08.07.2019	5.341	17.561	53,5
09.07.2019	5.593		42,4
10.07.2019	1.860		15,9
11.07.2019	2.364		18,2
12.07.2019	2.075		13,9
13.07.2019	328		2,6

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
14.07.2019			
15.07.2019	2.042	6.940	20,4
16.07.2019	1.094		9,3
17.07.2019	1.082		9,2
18.07.2019	1.804		13,1
19.07.2019	685		5,5
20.07.2019	233		1,6
21.07.2019			
22.07.2019	6.944	19.467	60,6
23.07.2019	1.020		8,5
24.07.2019	1.054		7,7
25.07.2019	3.886		32,5
26.07.2019	3.961		34,1
27.07.2019	2.602		23,4
28.07.2019			
29.07.2019	6.878	16.113	61,0
30.07.2019	1.260		11,9
31.07.2019	1.034		8,9
01.08.2019	1.363		9,8
02.08.2019	2.978		25,8
03.08.2019	2.600		23,8
04.08.2019			
05.08.2019	7.693	12.550	66,1
06.08.2019	1.149		9,9
07.08.2019	837		6,3
08.08.2019	1.452		10,1
09.08.2019	1.096		7,7
10.08.2019	323		2,3
11.08.2019			
12.08.2019	2.231	5.939	17,4
13.08.2019	896		7,6
14.08.2019	616		5,2
15.08.2019	1.354		11,5
16.08.2019	595		4,5
17.08.2019	247		1,8
18.08.2019			
19.08.2019	874	6.057	6,3
20.08.2019	485		4,3
21.08.2019	551		4,7
22.08.2019	1.341		9,6
23.08.2019	1.899		13,6

<b>Tag</b>	<b>Anzahl Datensätze</b>	<b>Anzahl Datensätze pro Woche</b>	<b>MB pro Datei</b>
24.08.2019	907		6,7
25.08.2019			
26.08.2019	898	6.072	6,7
27.08.2019	1.723		13,8
28.08.2019	978		8,5
29.08.2019	1.284		8,8
30.08.2019	658		4,9
31.08.2019	531		3,7
01.09.2019			
02.09.2019	2.231	8.870	16,0
03.09.2019	1.855		14,1
04.09.2019	1.102		8,0
05.09.2019	2.023		13,6
06.09.2019	1.187		8,4
07.09.2019	472		3,1
08.09.2019			
09.09.2019	992	8.214	7,1
10.09.2019	1.619		12,8
11.09.2019	1.460		10,6
12.09.2019	1.462		9,5
13.09.2019	873		6,7
14.09.2019	243		1,6
15.09.2019	1.565		22,5
16.09.2019	1.140	5.504	7,9
17.09.2019	981		7,6
18.09.2019	602		4,5
19.09.2019	1.448		9,6
20.09.2019	912		6,4
21.09.2019	421		2,9
22.09.2019			
23.09.2019	1.310	5.713	9,4
24.09.2019	1.047		8,0
25.09.2019	709		5,4
26.09.2019	1.608		11,0
27.09.2019	744		5,5
28.09.2019	295		2,1
29.09.2019			
30.09.2019	1.740		12,4

## Dokumentation

### D 1 Allgemeine Hinweise

Entwickelt für PHP 7.2.3, Firefox 60 und Internet Explorer 11.

Das Skript verwendet für Inserts in die Datenbank Transaktionen, die nicht alle SQL-Datenbanken unterstützen.<sup>83</sup>

Das Skript setzt die *maximum execution time* dynamisch auf 0. Falls dagegen Sicherheitsbedenken bestehen, kann in `const.php` ein Wert größer 0 gesetzt werden (s. Kap. 7.1).

Künftige Veränderungen im Format MARC 21-XML können sich auf das Skript auswirken und sollten daher beobachtet werden.<sup>84</sup>

Bei der Filterung von Datensätzen des Verlags GRIN, München, werden nicht-standardisierte Informationen aus HTML-Seiten des Verlags verwendet. Es muss von Zeit zu Zeit geprüft werden, ob das Verfahren noch korrekt arbeitet. (S. Kap. 6.5.)

### D 2 Installation

- Datenbank `reihe_o_tool` anlegen mit Kollation `utf8_unicode_ci`, `reihe_o_tool.sql` importieren
- Projektverzeichnis mit dem Namen `MATh` anlegen; `reihe_o_tool.zip` entpacken; den Inhalt des Ordners `MATh` im Projektverzeichnis speichern
- Zugangsdaten zur Datenbank setzen / ändern (in `accessData.php`)
- E-Mail-Adressen einrichten zum Versenden von Nachrichten durch das Skript sowie Empfänger-Adressen eintragen (in `email.php`)

---

<sup>83</sup> S. <https://www.php.net/manual/en/pdo.transactions.php> (zuletzt gesehen 6.12.2019).

<sup>84</sup> Z. B. über den Newsletter Metadatendienste der DNB, s. [https://www.dnb.de/DE/Professionell/Metatadendienste/Newsletter/newsletter\\_node.html#doc58228bodyText1](https://www.dnb.de/DE/Professionell/Metatadendienste/Newsletter/newsletter_node.html#doc58228bodyText1) (zuletzt gesehen 6.12.2019).

### **D 3 Benutzung**

Importdateien sollten regelmäßig gelöscht werden, um den Speicher zu entlasten. Auch Versanddateien und nicht mehr benötigte Datenbankinhalte müssen manuell gelöscht werden (jeweils über die Administrationsoberfläche).

#### Datenimport

Zum Import den Zeitraum eintragen, für den Daten importiert werden sollen. ‚Daten importieren‘ auswählen und auf ‚Absenden‘ klicken.

Es können Daten ab 1.1.2010 (Start der Reihe O) eingegeben werden.

Daten können für jeden Tag nur einmal abgerufen werden, mit Ausnahme von Tagen, für die der Import fehlerhaft war.

Zeiträume sollten nicht mehr als eine Woche betragen, um die Abrufzeit nicht zu lang werden zu lassen. Der Datenabruf für eine Woche kann bis zu etwa zwei Stunden dauern.

Das Dropdownmenü ‚Daten abgefragt für‘ informiert darüber, für welche Daten bereits ein Import erfolgreich ausgeführt wurde. Angezeigt werden auch Tage, für die keine Daten bei der DNB vorliegen. Tritt beim Import ein Fehler auf, wird der Tag nicht ins Menü eingetragen.

#### Pflichttitel auf Bestand prüfen

Den Zeitraum eintragen, für den Pflichttitel überprüft werden sollen, ‚Pflichttitel auf Bestand prüfen‘ auswählen und auf ‚Absenden‘ klicken.

Es können Titel nur für Tage überprüft werden, für die bereits ein Import erfolgt ist.

Die Bestandsprüfung kann beliebig oft wiederholt werden, solange die Titel nicht an das zuständige Team geschickt wurden.

Es wird jeder Titel am Bestand derjenigen Bibliothek abgeprüft, die nach aktuellem bayerischem Pflichtstückegesetz für ihn berechtigt ist. Derzeit (12.2019) ist das für alle Titel (auch) die BSB.

#### Pflichttitel ans Team schicken

Den Zeitraum eintragen, für den Titel verschickt werden sollen, ‚Pflichttitel ans Team schicken‘ auswählen und auf ‚Absenden‘ klicken.

Es können Titel nur für Tage verschickt werden, für die bereits ein Import erfolgt ist. Eine vorausgehende Bestandsprüfung ist nicht erforderlich, aber anzuraten.

Titel können nur einmal verschickt werden.

Die Titel werden nach Regierungsbezirken (Sitz des Verlags) auf die Sachbearbeiter\*innen aufgeteilt und separat per E-Mail verschickt.

#### Administrationsoberfläche

Hier können Änderungen an den Konfigurationsdateien vorgenommen und Importdateien gelöscht werden.

### **D 4 PHP-Dokumentation**

Die PHP-Dokumentation kann per [Pfad/zu/]MaTh/html/index.html aufgerufen werden.

### **D 5 Tests**

Die Mehrheit der Funktionen kann mit Komponententests (*unit tests*) getestet werden. Dazu wird das Testframework PHPUnit 8.4.1<sup>85</sup> von Sebastian Bergmann verwendet. Die Testcases liegen im Verzeichnis tests.

Die Tests können von der Kommandozeile aus ausgeführt werden. Dazu sind folgende Befehle einzugeben<sup>86</sup>:

#### Alle Tests ausführen

```
[Pfad/zu/]php.exe [Pfad/zu/]MATH/phpunit.phar --configuration  
[Pfad/zu/]MATH/config/phpunit.xml C:/xampp/htdocs/MATH/tests
```

#### Test für eine Klasse ausführen

```
[Pfad/zu/]php.exe [Pfad/zu/]MATH/phpunit.phar --configuration  
[Pfad/zu/]MATH/config/phpunit.xml Testklasse [Pfad/zu/Testklasse].php
```

---

<sup>85</sup> <https://phpunit.de/> (zuletzt gesehen 6.12.2019).

<sup>86</sup> Auf Windowssystemen sind die Slashes durch Backslashes zu ersetzen.

### Test für eine Methode ausführen

```
[Pfad/zu/]php.exe [Pfad/zu/]MATH/phpunit.phar --configuration  
[Pfad/zu/]MATH/config/phpunit.xml --filter "/(::Testmethode)(.*)?$/ " Testklasse  
[Pfad/zu/Testklasse].php
```

Die Tests können auch aus PHPStorm heraus ausgeführt werden. PHPStorm muss für die Ausführung von PHPUnit-Tests konfiguriert werden.<sup>87</sup>

### Alle Tests ausführen

Verzeichnis tests in der Dateieexplorerleiste markieren und den Test über die Menüpunkte ‚Run‘ - ‚Run...‘ aufrufen.

### Test für eine Klasse ausführen

Testklasse in der Dateieexplorerleiste markieren und den Test über die Menüpunkte ‚Run‘ - ‚Run...‘ aufrufen.

### Test für eine Methode ausführen

Testklasse öffnen, den Cursor in die gewünschte Methode stellen und den Test über die Menüpunkte ‚Run‘ - ‚Run...‘ aufrufen.

### Für Laufzeittests wird empfohlen, mit folgenden Tagen zu testen

- durchschnittliche Anzahl von Datensätzen für 1 Tag: 26.2.2019
- durchschnittliche Anzahl von Datensätzen für 1 Woche: 28.1. – 2.2.2019
- sehr niedrige Anzahl von Datensätzen für 1 Tag: 2.6.2019
- sehr hohe Anzahl von Datensätzen für 1 Tag: 1.12.2017

## **D 6 Inhalt der DVD**

Skriptcode des Reihe O-Tools (reihe\_o\_tool.zip)

Datenbankdump (reihe\_o\_tool.sql)

Text der Thesis (Balz\_Erwerbungsunterstuetzung.pdf)

---

<sup>87</sup> S. <https://www.jetbrains.com/help/phpstorm/using-phpunit-framework.html#> (zuletzt gesehen 13.12.2019).

## Literaturverzeichnis

[Abg-Bibl, 2009] Abgabe amtlicher Veröffentlichungen an Bibliotheken (Abgabe Bibliotheken – Abg-Bibl). Bekanntmachung der Bayerischen Staatsregierung vom 2. Dezember 2008 Az.: B II 2-480-30. In: Amtsblatt der Bayerischen Staatsministerien für Unterricht und Kultus und Wissenschaft, Forschung und Kunst, Nr. 1 (2009), S. 27-28. Online verfügbar unter <https://www.verkuendung-bayern.de/files/kwmbbl/2009/01/kwmbbl-2009-01.pdf> (zuletzt gesehen 6.12.2019).

[Arbeitstabelle MARC21, 1.7.2019] Arbeitstabelle Felddescription Titeldaten im Format MARC 21 (Stand: 1. Juli 2019). Verfügbar unter: [https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Metadatendienste/Exportformate/MARC21/marcFelddescriptionTitelExcel201903.zip?\\_\\_blob=publicationFile&v=2](https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Metadatendienste/Exportformate/MARC21/marcFelddescriptionTitelExcel201903.zip?__blob=publicationFile&v=2) (zuletzt gesehen 6.12.2019)

[dev.mysql.com] MySQL. The World's Most Popular Open Source Database. Developer Zone. Verfügbar unter: <https://dev.mysql.com> (zuletzt gesehen 6.12.2019).

[docs.phpdoc.org] PHPDocumentor. Documentation. Verfügbar unter: <https://docs.phpdoc.org> (zuletzt gesehen 6.12.2019).

[Edlich et al., 2011] Edlich, Stefan et al.: NoSQL. Einstieg in die Welt nichtrelationaler Web 2.0 Datenbanken. 2., akt. und erw. Auflage. München: Carl Hanser Verlag, 2011.

[Geirhos, 2018] Geirhos, Matthias: Entwurfsmuster: das umfassende Handbuch. 1. Auflage 2015 (2., korrigierter Nachdruck von 2018). Bonn: Rheinwerk (Galileo).

[Herold et al., 2017] Herold, Helmut et al.: Grundlagen der Informatik. 3., akt. Auflage. Hallbergmoos: Pearson, [2017].

[Klößner, 2015] Klößner, Kevin: Im Vergleich: NoSQL vs. relationale Datenbanksysteme. Verfügbar unter: [http://pi.informatik.uni-siegen.de/Mitarbeiter/mrindt/Lehre/Seminare/NoSQL/einreichungen/NOSQLTEC-2015\\_paper\\_9.pdf](http://pi.informatik.uni-siegen.de/Mitarbeiter/mrindt/Lehre/Seminare/NoSQL/einreichungen/NOSQLTEC-2015_paper_9.pdf) (zuletzt gesehen 6.12.2019).

[Lahres/Rayman, 2018] Lahres, Bernhard, Gregor Rayman: Objektorientierte Programmierung: das umfassende Handbuch. Bonn: Rheinwerk, 2018. 2., aktualisierte und erweiterte Auflage. Verfügbar unter: <http://openbook.rheinwerk-verlag.de/oop> (zuletzt gesehen 6.12.2019).

[MARC 21 Format] MARC 21 Format for Bibliographic Data. Library of Congress. Network Development and MARC Standards Office. 1999 Edition, Update No. 1 (October 2000) through Update No. 28 (May 2019). Verfügbar unter: <http://www.loc.gov/marc/bibliographic/> (zuletzt gesehen: 14.8.2019).

[Maurice, 2010] Maurice, Florence: PHP 5.3 & MySQL 5.1. Der Einstieg in die Programmierung dynamischer Websites. München: Addison-Wesley, 2010.

[Netzpublikationen in MARCXML, 17.7.2017] Netzpublikationen: Lieferung von Metadaten im Format MARCXML an die Deutsche Nationalbibliothek. Metadaten-Dokumentation für das Format MARCXML. Redaktionsteam Netzpublikationen

(Red.). Version 2.1 Stand: 17. Juli 2017. Verfügbar unter: <https://d-nb.info/1136935762/34>. (zuletzt gesehen: 6.12.2019).

[OAI-PMH, 2005] Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14, Document Version 2005/05/03T22:51:00Z. Verfügbar unter: <https://www.openarchives.org/OAI/2.0/guidelines.htm> (zuletzt gesehen: 6.12.2019).

[Oberfell/Thalhofer, 2018] Oberfell, Jörg, Anja Thalhofer: Mit der Wühlmaus „Anawülas“ unterwegs – Das neue Auswertesystem der WLB für Titeldaten. In: WLBforum (2018) 1, S. 7-19. Verfügbar unter: [https://www.wlb-stuttgart.de/fileadmin/user\\_upload/die\\_wlb/WLB-Forum/WLBforum\\_2018\\_1\\_WEB.pdf](https://www.wlb-stuttgart.de/fileadmin/user_upload/die_wlb/WLB-Forum/WLBforum_2018_1_WEB.pdf) (zuletzt gesehen 6.12.2019).

[Oberfell/Thalhofer, 2019] Oberfell, Jörg, Anja Thalhofer: Die Wühlmaus wühlt weiter... Mit Anawülas Geschäftsgänge digitalisieren, landesbibliothekarische Aufgaben erfüllen, Quellen lieferantenunabhängig auswerten und Medien erwerben. In: WLBforum (2018) 1, S. 7-17. Verfügbar unter: [https://www.wlb-stuttgart.de/fileadmin/user\\_upload/die\\_wlb/WLB-Forum/WLBforum\\_2019\\_1\\_WEB.pdf](https://www.wlb-stuttgart.de/fileadmin/user_upload/die_wlb/WLB-Forum/WLBforum_2019_1_WEB.pdf) (zuletzt gesehen 6.12.2019).

[php] php. Verfügbar unter: [www.php.net](http://www.php.net) (zuletzt gesehen 6.12.2019).

[PflStG, 1986] Gesetz über die Ablieferung von Pflichtstücken. Pflichtstückegesetz (PflStG). Vom 6. August 1986. Verfügbar unter: [https://www.bsb-muenchen.de/fileadmin/pdf/pflichtablieferung/pflichtstueckegesetz\\_200602.pdf](https://www.bsb-muenchen.de/fileadmin/pdf/pflichtablieferung/pflichtstueckegesetz_200602.pdf) (zuletzt gesehen 6.12.2019).

[Pflichtstückesammlung, 1986] Sammlung der Pflichtstücke nach dem Gesetz über die Ablieferung von Pflichtstücken. Bekanntmachung des Bayerischen Staatsministeriums für Unterricht und Kultus vom 11. November 1986 Nr. V/2 - K 3474/2 - 12/148282. In: Amtsblatt der Bayerischen Staatsministerien für Unterricht und Kultus und für Wissenschaft und Kunst (KWMBL.) I 11986, Nr. 24/1986, S. 530.

[phpunit] Sebastian Bergmann: PHPUnit Manual. Edition for PHPUnit 8.4. Updated on Nov 09, 2019. Verfügbar unter <https://phpunit.readthedocs.io/en/8.4/> (zuletzt gesehen 6.12.2019).

[Pierer, 2012] Pierer, Christian: Literaturverwaltung einmal anders - Optimierungspotentiale im Erwerbungsprozess, verfügbar unter [https://opus4.kobv.de/opus4-bib-info/files/1070/Pierer\\_Bib\\_Tag.pdf](https://opus4.kobv.de/opus4-bib-info/files/1070/Pierer_Bib_Tag.pdf) (zuletzt gesehen 6.12.2019).

[Smith/Sasse, 2010] Smith, Patricia A., Margo Sasse: Automated Acquisitions. In: Encyclopedia of library and information sciences. Marcia J. Bates et al. (Hg.). Boca Raton: CRC Press, 2010. 3. ed. Vol. 1, S. 420-429.

[Schlossnagle, 2006] Schlossnagle, George: Professionelle PHP 5-Programmierung. Entwicklerleitfaden für große Webprojekte mit PHP 5. München: Addison-Wesley, 2006.

[smarty.net] smarty™ template engine. New Digital Group, 2019. Verfügbar unter: <https://www.smarty.net> (zuletzt gesehen 6.12.2019).

[SRS IEEE 830-1998] IEEE Recommended Practice for Software Requirements Specifications. Approved 25 June 1998 IEEE-SA Standards Board, New York: The Institute of Electrical and Electronics Engineers 1998. Verfügbar unter <http://www.math.uaa.alaska.edu/~afkjm/cs401/IEEE830.pdf> (zuletzt gesehen 6.12.2019).

[Stauffer, 2019] Stauffer, Matt, Laravel Up & Running. A Framework for Building Modern PHP Apps, Sebastopol, CA: O'Reilly, 2019, 2. Auflage.

[Stokes, 2015] Stokes, John M.: XML parsing with PHP. Alexandria, VA; muske-teers.me, LLC; April 2015.

[Strauch/Rehm, 2007] Strauch, Dietmar; Margarete Rehm: Lexikon Buch, Bibliothek, Neue Medien. 2., aktualisierte und erw. Ausg. München: Saur, 2007.

[TB Informatik] Schneider, Uwe (Hrsg.): Taschenbuch der Informatik: mit 99 Tabellen, 7., neu bearb. Auflage, München: Fachbuchverlag Leipzig im Carl-Hanser-Verlag, 2012.

[Theis, 2018] Theis, Thomas: Einstieg in PHP 7 und MySQL, Bonn: Rheinwerk Verlag, 2018. 13., aktualisierte Auflage.

[w3schools.com] w3.schools.com. The World's Largest Web Developer Site. Verfügbar unter: <https://www.w3schools.com> (zuletzt gesehen 6.12.2019).

[Wiegers, 1990] Wiegers, Karl. E., [Template for a Software Requirements Specification]. Verfügbar unter: <https://goo.gl/nsUFwy> (zuletzt gesehen 6.12.2019).

[WLB-Forum 2/2019] WLB-Forum 2/2019. Verfügbar unter: [https://www.wlb-stuttgart.de/fileadmin/user\\_upload/die\\_wlb/WLB-Forum/WLBforum\\_2019\\_2\\_WEB.pdf](https://www.wlb-stuttgart.de/fileadmin/user_upload/die_wlb/WLB-Forum/WLBforum_2019_2_WEB.pdf) (zuletzt gesehen 6.12.2019).

## Abbildungsverzeichnis

Abb. 1. Suchstrings des ABM-Tools in CQL-Syntax	11
Abb. 2. Die Oberfläche des ABM-Tools	12
Abb. 3. Import (Aktivitätsdiagramm)	29
Abb. 4. Bestandsprüfung (Aktivitätsdiagramm)	32
Abb. 5. Kauf (Aktivitätsdiagramm)	34
Abb. 6. Datenexport (Aktivitätsdiagramm)	36
Abb. 7. Klassendiagramm	42
Abb. 8. Controller	43
Abb. 9. Model	43
Abb. 10. Data (verkürzt dargestellt)	44
Abb. 11. Database	44
Abb. 12. DataTransformation	45
Abb. 13. FileHandler	45
Abb. 14. Holdings	46
Abb. 15. Import	46
Abb. 16. Mail	47
Abb. 17. Relevance	47
Abb. 18. Validation	48
Abb. 19. XMLProcessor	48

## **Selbständigkeitserklärung**

Ich versichere, dass die vorliegende Arbeit von mir selbständig und ohne unerlaubte Hilfe angefertigt worden ist. Ich habe alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen sind, durch Zitate bzw. Literaturhinweise als solche kenntlich gemacht.

Ort, Datum

---

(Unterschrift)