

Masterarbeit

zur Erlangung des akademischen Grades
Master

Technische Hochschule Wildau

Fachbereich Wirtschaft, Informatik, Recht

Studiengang Bibliotheks-informatik (M. Sc.)

Thema (deutsch): Konzeption und Entwicklung eines datengetriebenen Unterstützungssystems für Etatplanung und Mittelallokation einer hybriden Spezialbibliothek

Thema (englisch): Design and development of a data-driven support system for budget planning and resource allocation of a hybrid library

Autor/in: Peter Breternitz

Seminargruppe: BIM/17

Betreuer/in: Dipl.-Informatiker Sascha Szott

Zweitgutachter/in: Dr. Frank Seeliger

Eingereicht am: 10.03.2021

Konzeption und Entwicklung eines datengetriebenen
Unterstützungssystems für Etatplanung und Mittelallokation einer
hybriden Spezialbibliothek

von

Peter Breternitz

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

ZUSAMMENFASSUNG

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

INHALTSVERZEICHNIS

1	EINFÜHRUNG	1
1.1	Problemstellung	1
1.2	Ziel der Arbeit	3
1.3	Verwandte Arbeiten	5
1.4	Gliederung der Arbeit	7
2	THEORETISCHE GRUNDLAGEN	9
2.1	Bibliothek und Statistik	9
2.2	Datenvisualisierung	13
2.3	Business-Intelligence-Systeme	17
3	AUSGANGSSITUATION	23
3.1	Bibliothek	23
3.1.1	Allgemeines	23
3.1.2	Organisatorische Einbettung	24
3.1.3	Informationsdienstleistungen	24
3.1.4	Evaluation der Informationsdienstleistungen	25
4	KONZEPTION EINER LÖSUNG	29
4.1	Anforderungsanalyse	29
4.1.1	Vision und Ziele	30
4.1.2	Rahmenbedingungen	30

4.1.3	Funktionale Anforderungen	31
4.1.4	Nicht-funktionale Anforderungen	33
4.1.5	Anwendungsfälle	34
5	DISKUSSION DER UMSETZUNG	41
5.1	Implementierung	41
5.1.1	Technische Details der Implementierung	41
5.1.2	Systemarchitektur	44
5.1.3	Teilsysteme	45
5.2	Funktionsweise	58
5.2.1	Daten-Import	58
5.2.2	Dashboard	58
5.3	Bewertung	60
5.3.1	Umgesetzte Anforderungen	60
6	SCHLUSS	61
6.1	Soll-Ist-Vergleich (Stand der Umsetzung)	61
6.2	Lessons learned	61
6.3	Welche Themen wurden nicht bearbeitet	61
6.4	Welche Themen sind im Anschluss denkbar	61
	TABELLENVERZEICHNIS	63
	ABBILDUNGSVERZEICHNIS	65
	QUELLCODEVERZEICHNIS	67
	AKRONYME	69
	LITERATURVERZEICHNIS	71

1 EINFÜHRUNG

Als Antwort auf die anhaltende Unsicherheit im Bereich der öffentlichen Gesundheit, die die Ausbreitung der COVID-19-Pandemie hervorgerufen hat, entwickelte zu Beginn des Jahres 2020 ein Team um die Professorin Lauren Gardner der *Johns Hopkins University* ein Dashboard. Dieses visualisiert die gemeldeten Fälle der COVID-19-Pandemie weltweit. Das Dashboard wurde entwickelt, um Forschenden, Gesundheitsbehörden und der breiten Öffentlichkeit ein benutzerfreundliches Instrument an die Hand zu geben, mit dem sich der Ausbruch leicht verfolgen lässt. Visualisiert durch eine Weltkarte und unterschiedlich großen Punktmarkierungen zeigt das Dashboard die gegenwärtige Ausbreitung. Zusätzlich werden Zahlen der bestätigten COVID-19-Fälle, der Todesfälle und der Genesungen für alle betroffenen Länder angezeigt[Vgl. [DDG20](#), S. 533]. In der Anfangszeit des Dashboards wurden die Daten noch manuell gesammelt und bearbeitet. Mittlerweile wurde ein halb-automatischer Prozess implementiert. Durch die übersichtliche Darstellung auf dem Dashboard können rasch Informationen über die Gefahrenlage von einer breiten Öffentlichkeit rezipiert werden. Daraus kann ein Verständnis der derzeitigen Maßnahmen bei der Öffentlichkeit entstehen sowie die Informationen in eigene Handlungsoptionen integriert werden.

1.1 PROBLEMSTELLUNG

Ausgehend von ökonomischen, informationstechnologischen und marktpolitischen Veränderungen in den vergangenen Jahrzehnten, sind Bibliotheken dazu veranlasst, ihr Budget

hinsichtlich der Informationsbedarfe ihrer Nutzer:innen behutsamer zu planen und sich in zunehmenden Maße gegenüber ihren Unterhaltsträgern zu rechtfertigen. Die Relevanz von bibliothekarischen Statistiken ist in diesem Zusammenhang größer geworden. Deswegen ist es wichtig, Daten aus den bibliothekarischen Bereichen zu aggregieren, statistisch zu erheben und auszuwerten, um auf Basis der daraus erzielten Erkenntnisse handeln zu können. Die Transparenz von statistischen Daten sorgt für eine bessere Grundlage in den Verhandlungen mit den Stakeholdern einer Bibliothek. Zudem wird durch sie der Einsatz des Bibliotheksbudgets zielgerichteter auf die Bedürfnisse der Nutzer:innen zugeschnitten. Dazu ist es zweckmäßig, alle anfallenden Daten für die Budgetplanung und Mittelallokation einer Bibliothek zentral zu sammeln und mit geeigneten statistischen Methoden und Verfahren langfristig auszuwerten. Um den Wert dieser aus den Daten gewonnenen Information elegant den Stakeholdern zu kommunizieren und zu präsentieren, können geeignete Verfahren der Datenvisualisierung zum Einsatz kommen. Die technische Realisierung kann durch gewöhnliche Tabellenkalkulationsprogramme erfolgen. Um den mitunter hohen Zeitaufwand und die Fehleranfälligkeit manueller Prozesse einerseits zu minimieren und den Automatisierungsgrad hinsichtlich der Aggregation und Auswertung der bibliothekarischen Daten andererseits zu erhöhen, können aber auch andere technische Umsetzungen eingesetzt werden. In Bereichen der Wirtschaft kommen Business Intelligence (BI)-Systeme zum Einsatz, die IT-basiert Entscheidungsfindungen unterstützen (siehe dazu [Abschnitt 2.3](#)).

Es gibt bereits eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf *BI-Systemen* basieren. Zu nennen wären *AlmaAnalytics* für das Next-Generation-Library-System *Alma* von *ExLibris* [[Lib20b](#)], *BibControl* vom Online Computer Library Center (OCLC) [[OCL20](#)], *CollectionHq* von *Baker & Taylor* [[BT20](#)] oder *Libinsight* von *SpringShare* [[Spr20](#)]. Darüber hinaus gibt es *Business Intelligence-Systeme*, die von Bibliotheken für Reporting, Datenanalyse und Datenvisualisierung adaptiert werden, wie zum Beispiel *Tableau* von der Firma *Tableau Software* [[Sof20](#)], *Crystal Reports* von *SAP*

[SAP20] oder *Microsoft BI* von *Microsoft* [Mic20]. Diese Applikationen sind entweder an bestimmte Bibliothekssysteme zurückgebunden, limitiert in ihren Funktionen oder zu generisch.

Für die Spezialbibliothek des Max-Planck-Institut für empirische Ästhetik (MPI EA) begründet sich die Notwendigkeit für eine an *BI-Systemen* angelehnte Applikation durch das Fehlen eines zentralen Nachweissystems für bibliothekarische Statistiken in der Bibliothek. Überdies wird sowohl vom Hessisches Bibliotheksinformationssystem (HeBis)-Verbund, deren Mitglied die Bibliothek ist, als auch von der max-planck-digital-library (mpdl) keine Systeme in dieser Richtung angeboten. Ebenso ist ungewiss, wann die Ablösung des schon betagten Bibliothekssystems hin zu einem neuen Next-Generation-Library-System in *HeBis* stattfinden wird und ob es ein Modul zur statistischen Datenerhebung mitbringen wird. Des Weiteren ist das Erfordernis, bibliothekarische Geschäftsprozesse zu evaluieren und die Servicedienstleistungen bezüglich der Ziele der Institution noch weiter zu optimieren, von großer Relevanz. Die zu entstehende Applikation könnte hierbei helfen, systematisches Controlling einzuführen und das Bibliotheksmanagement weiter zu professionalisieren. Mit dem Ende der Konsolidierungsphase der Bibliothek, die im Zuge des *MPI EA* 2013 gegründet wurde, tritt sie ein in eine Phase, in der ab dem Jahr 2021 Budgetplanungen eine größere Rolle spielen werden.

1.2 ZIEL DER ARBEIT

Das Ziel der Arbeit ist die Schaffung eines Dashboards für die Budgetplanung in der Spezialbibliothek des *MPI EA*. In Anlehnung an *BI-Systeme* soll ein ganzheitliches System als Proof-of-Concept entstehen, mit dem systematisch die relevanten Daten der hybriden Spezialbibliothek aggregiert, statistisch mit geeigneten und modernen Datenvisualisierungen analysiert und ausgewertet werden sollen. Darüber hinaus soll es möglich sein, aus dem System ausgewählte Resultate automatisiert als Standardbericht zu exportieren, um

1 Einführung

diese als *factsheet* gegenüber Stakeholdern der Bibliothek präsentieren zu können. Um künftigen Anforderungen gewachsen zu sein, soll das Dashboard modulbasiert programmiert werden und dadurch leicht erweiterbar sowie eventuell von anderen Bibliotheken nachnutzbar sein.

Es gibt einen wachsenden Markt für Systeme, die solche Anwendungen möglich machen. Dieser wächst im Schatten von Data-Science und Big Data. Der Markt verfügt über ausgereifte und mächtige Frameworks, die statistische Auswertungen mit wenig Programmieraufwand erlauben. Der höhere zeitliche Aufwand bei solchen statistischen Auswertungen liegt demgegenüber einerseits in der Aufbereitung der Daten, insbesondere dann, wenn die Daten aus heterogenen Datenquellen kommen. Das setzt dementsprechend eine genaue Analyse der Daten voraus. Andererseits kann die Zusammenführung der einzelnen statistischen Auswertungen in eine Applikation eine nicht leicht zu nehmende Hürde darstellen. Die Entwicklung von interaktiven Dashboards oder ähnlichen Web-Anwendungen ist ebenfalls leichter geworden, da Kenntnisse in einer Programmiersprache durchaus ausreichend sind, um eine solche Applikation zu programmieren.

Folgende Aufgaben und Zwecke soll das System lösen können: Die Daten sollen aus den heterogenen Datenquellen mit einem Extract, Transform, Load (ETL)-Prozess bearbeitet und in geeigneter Form gespeichert und bereitgestellt werden. Die Auswertungen erfolgen mittels statistischer Verfahren, die aus den Formulierungen der Fragen an die Daten abgeleitet werden. Dabei sollen Verfahren der deskriptiven und explorativen Statistik zum Einsatz kommen. Die Auswertungen werden dann über das interaktive Dashboard visuell dargestellt. Die visuelle Darstellung der Daten soll nach einstellbaren Parametern schrittweise verfeinert werden können.

Ein wesentlicher Zweck des Dashboards ist das Monitoring der laufenden Budgetkosten für die Medien. So sollen die Kosten für das laufende Jahr auch im Vergleich zu den Vorjahren dargestellt werden können. Ein anderer wichtiger Baustein sind die Ausleihzahlen der Medien. Diese können anhand der Dimensionen der Zeit oder der Aufstellungssystematik

analysiert werden. Diese Analysen können sich nach Zeitraum beziehungsweise nach Bestandssegmenten bis hinunter auf Titelebene als kleinste Ebene ausdifferenzieren. Ebenso vorgesehen ist eine Analyse der Neuerwerbungen nach Zeit anhand der Dimension der Aufstellungssystematik.

Der Standardbericht greift auf vorberechnete Auswertungen und Darstellungen zurück und generiert anfrageorientiert einen Bericht, der die wichtigsten Key Performance Indicators (KPI) der Bibliothek enthält. Er wird im Format PDF verteilt und kann ohne bibliothekarisches Domänenwissen gelesen werden.

Das Dashboard ist für das Monitoring und die Analyse der Daten durch die Bibliothek gedacht. So kann die Bibliotheksleitung die Ausgaben für Medien nach Lieferanten überwachen und gegebenenfalls steuernd eingreifen. Anhand der Analyse der Daten kann die Bibliotheksleitung das Jahresbudget besser steuern. Am Wachstum des Bestandes beziehungsweise der Bestandssegmente können die Bibliotheksmitarbeiter:innen ablesen, welche Bestandsgruppe wie schnell wächst. Der Standardbericht als aggregierte Form wichtigster *KPI* ist für die Kommunikation und Präsentation gegenüber der Institutsleitung gedacht.

1.3 VERWANDTE ARBEITEN

Es gab in den letzten Jahren Versuche an Universitätsbibliotheken Dashboards zu entwickeln, um daten-getriebene Entscheidungen zu unterstützen. So wurde an den *New York University Health Sciences Libraries* ein Dashboard entwickelt, das versucht, möglichst viele Metriken aus bibliothekarischen Dienstleistungen aufzunehmen. Die Architektur des Dashboards besteht aus drei Hauptteilen. Die Daten werden mit Import-Skripten aus den verschiedenen Datenquellen bezogen, mit einem *ETL-Prozess* bearbeitet und in ein Data Warehouse geladen. Das Data Warehouse stellt eine einfache MySQL-Datenbank dar. Die Daten werden aus dieser mit einem Mix von PHP/Javascript-Skripten in einem

1 Einführung

Dashboard mit unterschiedlichen Diagrammen dargestellt [MH12]. Ebenfalls wurde an der *James C. Kirkpatrick Library der University of Central Missouri* ein Dashboard zur Unterstützung Evidenz-basierter Entscheidungen der Bibliotheksleitung entwickelt. Vom Entwicklungs-Team wurde sich gegen kommerzielle Lösungen entschieden und ein Dashboard mit dem Ruby-on-Rails-Framework entwickelt. Es zeigt verschiedene Daten wie Study-Room- oder Computer-Usage-Data [Lib20a]. Horne-Popp, Tessone und Welker geben detaillierten Einblick in die Designentscheidungen und Problematiken während der Entwicklung dieses Tools [vgl. HTW18, 194 ff.].

An der *Universitätsbibliothek der Technischen Universität Berlin* wird *Alma Analytics* benutzt, was eine Business-Intelligence-Lösung des Bibliothekssystemanbieters *ExLibris* darstellt. Golas beschreibt die Neuimplementierung der statistischen Abfragen nach Einführung des cloudbasierten Bibliotheksmanagementsystems *Alma* an der Universitätsbibliothek. Neu implementiert wurden unter anderem Trends nach der Regensburger Verbundklassifikation (RVK), die RVK-Schlagwörter-Zuordnung, Statistiken über Ausleihen und Vormerkungen [Gol18].

In eine ähnliche Richtung wie das zu entwickelnde Tool, geht die an der *Technischen Hochschule Wildau* entwickelte Software *BibloVis*. *BibloVis* basiert auf einem modular aufgebauten Client-Server-Modell und ermöglicht die Einbindung und Visualisierung von Daten wie der Katalognutzung, Raumauslastungen und anderen bibliothekarischen Service-Angeboten. Die Grundlage bildet die Auswertungen von CSV- und XML-Dateien [BS15].

Der Einsatz von *Tableau* an den *Ohio State University Libraries* wird von Murphy beschrieben. Anhand von zwei Projekten werden die Einsatzmöglichkeiten insbesondere der Datenvisualisierungsoptionen dieser Software dargestellt. Unter anderem wurden die Transaktionsprotokolldateien auf Bibliotheksleitfäden untersucht, die von der Bibliothek zu den Universitätskursen herausgegeben wurden. Diese wurden über einen Zeitraum von 2009-2012 ausgewertet und mit verschiedenen Datenvisualisierungen dargestellt [vgl. Mur13, 469 f.].

Ein gutes Beispiel für ein datengetriebenes Unterstützungssystem findet sich bei Spielberg, der sich mit dem Thema der Bestandspflege an der *Universitätsbibliothek Essen* befasst und eine Applikation (weiter-)entwickelt hat, die die Fachreferent:innen bei der Aussonderung und Erwerbung von Medien unterstützt [Spi17].

Ebenso finden sich in der Fachliteratur Ansätze, die vorrangig anhand einzelner Fragestellungen hinsichtlich der Bestandsentwicklung[Hug16] oder anderer bibliothekarischer Servicedienstleistungen[KM20; KWC06; Mey18] verschiedene statistische Analysen vollzogen und diese visualisiert haben.

Fast alle Projekte sind an größeren Bibliotheken mit ganz unterschiedlichen softwaretechnischen Herangehensweisen[FF16; WH13] und Zielen[Phe12] entstanden.

Zudem fehlt ein Beispiel in der Literatur, das holistisch alle relevanten Daten, die in den Geschäftsgängen und Servicedienstleistungen insbesondere einer Spezialbibliothek entstehen, aggregiert, auf diesen Daten automatisch statistische Analysen ausführt und diese mit modernen Visualisierungstechniken interaktiv darstellt.

1.4 GLIEDERUNG DER ARBEIT

Im Kapitel 2 werden die theoretischen Grundlagen für die folgenden Kapitel gelegt. Das Kapitel beschreibt den Rahmen, in dem die Entwicklung eines Dashboards eingebettet ist. Dabei wird herausgestellt, wie wichtig Statistik im bibliothekarischen Bereich ist, was Datenvisualisierungen sind und warum sie eingesetzt werden sollen und welche Anleihen *BI-Systeme* für das zu entstehende System liefern können. Im Kapitel 3 wird die Bibliothek vorgestellt und darauf eingegangen, welche bibliothekarischen Statistiken bereits erhoben wurden. Nachdem die Ausgangssituation bestimmt wurde, wird mit der Anforderungsanalyse im Kapitel 4 das generierte Wissen von Kapitel 2 aufgegriffen und die Konzeption einer Lösung vorgestellt. Im Kapitel 5 wird die Umsetzung diskutiert. Bevor das System bewertet wird, wird das Design und die Implementierung vorgestellt. Das Fazit wird im

1 Einführung

Kapitel 6 mit dem Stand der Umsetzung, den *lessons learned* und einem Ausblick auf Themen, die noch bearbeitet werden könnten, gezogen.

2 THEORETISCHE GRUNDLAGEN

In diesem Kapitel wird der theoretische Rahmen für die weiteren Kapitel gelegt. Im ersten Abschnitt werden die Grundlagen der Budgetplanung und Mittelallokation im Zusammenhang mit bibliothekarischen Statistiken erläutert. Der darauffolgende Abschnitt handelt von Datenvisualisierungen und deren Einsatz für Datenrepräsentationen und Datenpräsentation. Abschließend wird das Modell der *Business Intelligence-Systeme* als dritter Baustein zur Lösung der Problemstellung dieser Arbeit vorgestellt.

2.1 BIBLIOTHEK UND STATISTIK

Die Etatplanung von Bibliotheken richtet sich nach deren Informations- und Versorgungsauftrag. Seit Beginn der 1990er Jahre müssen sich Bibliotheken mit den Auswirkungen einer veränderten Medienlandschaft auseinandersetzen. Sie kämpfen mit einem größer werdenden Informationsangebot, den steigenden Preisen auf dem Publikationsmarkt, den zunehmenden Kommerzialisierungstendenzen in der Verlagslandschaft und neuen Medientypen. Zu nennen wären hier konkret: die Explosion der Zeitschriftenpreise im Bereich der Science, Technology, and Medicine (STM), die Konzentration auf wenige Verlage, und dem Aufkommen von ebooks. Demgegenüber steigen Bibliotheksetats nur mäßig. Demzufolge geht ein Kaufkraftverlust der Bibliotheken einher [vgl. [Mor15](#), 164 ff.]. Bibliotheken haben gegen diesen Trend Instrumente entwickelt, um den Informationsauftrag trotz dieser Widrigkeiten zu erfüllen. So entstehen seit Mitte der 1990er Jahre von Bund und Ländern geförderte Konsortien, um den Kostendruck auf Bibliotheken insbe-

sondere im Bereich der elektronischen Fachinformationen zu mildern. Neue Geschäftsmodelle werden zur Abfederung der Kosten entwickelt, um Preisnachlässe bei großen Verlagen zu erzielen [vgl. [Mor15](#), 169 ff.]. Das Projekt *Deal* – ein Projekt der Hochschulrektorenkonferenz (HRK) in Zusammenarbeit mit den wissenschaftlichen Einrichtungen in Deutschland – konnte so in den vergangenen Jahren Verträge mit den Verlagen *Springer* und *Wiley* erfolgreich aushandeln [vgl. [Dea20](#)].

Um den Veränderungen des Publikationsmarktes lokal in der Bibliothek zu begegnen, wird es immer wichtiger, das Bibliotheksbudget und die Mittelallokation kosteneffizient zu planen. Dies geschieht bisher in größeren Bibliotheken durch Etatbedarfs- und Etatverteilungsmodelle [vgl. [Mor15](#), 172 ff.]. Diese Modelle basieren auf der statistischen Erhebung von bibliothekarischen Kennzahlen.

Bibliotheksstatistik reflektiert das Gestern, Heute und Morgen, indem sie die bibliothekarischen Servicedienstleistungen evaluiert und den zukünftigen Zielen und Aufgaben anpasst [vgl. [Jil04](#), 2 f. vgl. [Lai13](#), S. 462]. Im deutschen Bibliothekswesen gibt es die umfangreiche Deutsche Bibliotheksstatistik (DBS). Träger der *DBS* sind das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz), das Kompetenznetzwerk für Bibliotheken (KBN), die Kultusministerkonferenz (KMK) sowie die teilnehmenden Bibliotheken. Aufgabe der *DBS* ist die jährliche statistische Datenerhebung von Bibliothekskennzahlen. Seit 1999 werden die Daten nur noch online erfasst, ausgewertet und präsentiert [vgl. [SB08](#), S. 2]. Daneben gab es den Bibliotheksindex (BIX), der ursprünglich für die Leistungsmessung in Öffentlichen und Wissenschaftlichen Bibliotheken konzipiert wurde. Dieser wurde 2015 aber aufgrund von Finanzierungsproblemen eingestellt.

Bibliothekskarische Kennzahlen werden durch quantitative und qualitative Evaluationsverfahren erhoben. Diese Verfahren sind auf den Bestand der Bibliothek zentriert. Bestand ist nach Johannsen und Mittermaier „... die Gesamtheit aller Medien, die eine Bibliothek ihren Nutzern anbietet, sei es, dass sie diese 'physisch' besitzt, sei es, dass sie entsprechende Nutzungsrechte erworben hat.“ [[JM15](#), S. 252] Als Typen der Bestandsevaluation sind

sammelungs-, nutzungsbezogene und nutzer:innenbezogene Evaluationen zu nennen [vgl. Joh14, S. 302]. Basiert die sammlungs- und nutzungsbezogene Evaluation auf quantitativen Daten, greift die nutzer:innenbezogene Evaluation zumeist auf qualitative Daten zurück [vgl. BS04, 461 ff.].

Die sammlungsbezogene Evaluation betrifft die Größe des Bestandes und das Wachstum über die Jahre. Die Bestimmung der Bestandsstärke- und tiefe, der Ausgewogenheit in den Bestandssegmenten sind Ziele der sammlungsbezogenen Evaluation. Ebenfalls lässt sich die Frage nach der aktuellsten Literatur im Bestand oder in einem Segment durch die sammlungsbezogene Evaluation klären [vgl. Lyo10, 48 f.].

Nutzungsbezogene Evaluation umfasst die Lesesaalnutzung, die Ausleihe vor-Ort, die Nutzung des Fernleihservices oder der Dokumentenlieferdienste sowie die Online-Nutzung von elektronischen Ressourcen [vgl. JM15, 254 ff.]. Die Frage nach den Zugriffsstatistiken auf elektronischen Ressourcen beansprucht in der nutzungsbezogenen Evaluation einen größer werdenden Raum. Die internationale Organisation *Counting Online Usage of Networked Electronic Resource* (COUNTER) gibt dazu die COUNTER-Statistiken heraus. Mitglieder der Organisation sind Verlage, Bibliotheken und Zwischenhändler. Die COUNTER-Statistiken sind mittlerweile der Quasi-Standard für die Zugriffsstatistiken auf elektronische Ressourcen geworden. Diese werden getrennt nach Art der Informationsressourcen in verschiedenen Reports herausgegeben. [vgl. JM15, 260 ff.]. Mittlerweile ist die fünfte Iteration der COUNTER-Statistiken COP 5 erschienen [vgl. Cou20]. Die Bibliotheken sind bei dem Bezug von diesen Statistiken auf die Unterstützung der Verlage angewiesen. Diese stellen unregelmäßig die COP 5-Statistiken zur Verfügung.

Ziele der nutzungsbezogenen Evaluation sind die Identifizierung von ausleihträchtigen Medienbeständen (Vormerkungs- und Rennerlisten) und die De-Akquisition schlecht oder gar nicht genutzter Titel. Ebenso kann die Evaluation von Fernleih- und Dokumentenlieferungen Hinweise auf Bestandslücken liefern. Als Konsequenz aus den COUNTER-Statistiken kann die Abbestellung von elektronischen Ressourcen resultieren.

2 Theoretische Grundlagen

Die nutzer:innenbezogene Evaluation ist auf den Nutzer:innenkreis der Bibliothek und deren Informationsbedürfnisse zentriert. Nutzer:innenbezogene Evaluation benutzt qualitative Daten, die sie aus Befragungen erhebt [vgl. JM15, 255 ff. vgl. Joh14, S. 302].

Die einzelnen Evaluationen vermitteln ein umfassendes Gesamtbild der Bibliothek und deren Service-Dienstleistungen. Die datengetriebenen Evaluationsauswertungen bieten Hinweise auf Optimierungen der bibliothekarischen Service-Dienstleistungen. Die Auswertungen können durch die Bibliotheksleitung aufgenommen werden und in strategische (zukünftige) Entscheidungen einfließen. So kann ein detailliertes Erwerbungsprofil und somit eine gezieltere Erwerbungspolitik entstehen. Dadurch wird das Management der Ressourcen effektiver und effizienter [vgl. Joh14, S. 297]. Gegenüber Stakeholdern kann auf der Grundlage der Evaluationen gezielt um Budget verhandelt werden.

The purpose of the statistics is to give the management of the library or another decision-maker a satisfactory and correct picture about the situation of the library as a support to them - the statistics are the mirror of the library! [Lai13, S. 463]

Um ein ansprechendes und korrektes Bild der Situation der Bibliothek zu präsentieren, helfen sorgsam ausgewählte Datenvisualisierungen.

2.2 DATENVISUALISIERUNG

Datenvisualisierungen sind wirkmächtig. Sie stellen einen Weg dar, statistische Informationen effizient zu kommunizieren [vgl. Tuf19, S. 15], indem sie Daten mit visuellen Reizen ausstatten, die vom menschlichen Auge aufgenommen und vom menschlichen Gehirn schnell verarbeitet werden können [vgl. Few09, S. 32]. Zusammenhänge, Trends und Ausnahmen einer großen Datenmenge sind in einer Zahlenkolonne schwieriger zu entdecken als mit einer geeigneten Datenvisualisierung. Datenvisualisierungen ermöglichen den visuellen Vergleich von verschiedenen Informationen. Sie können eine große Anzahl von Datensätzen kompakt darstellen. Datenvisualisierungen können nicht nur die Informationen von verschiedenen Blickwinkeln anzeigen, sondern die Informationen auch mit unterschiedlicher Granularität darstellen [vgl. ML13, S. 245]. Visualisierungen benötigen Daten. Daten benötigen Visualisierungen, um ihren Wert besser präsentieren zu können [vgl. Kir19, S. 16].

Im Folgenden werden der inhaltliche Bezugsrahmen und die Merkmale von Datenvisualisierungen näher erläutert. Datenvisualisierungen sind Verfahren der deskriptiven und explorativen Statistik beziehungsweise der explorativen Datenanalyse. Im Allgemeinen bilden sowohl die deskriptive Statistik als auch die explorative Datenanalyse keine Hypothesen. Beide treffen nur Aussagen zu vorliegenden Datensätzen. Dennoch gibt die explorative Statistik Hinweise für eine mögliche Hypothesenbildung in der weiterführenden Analyse. Die Datenvisualisierung hat sich aus den explorativen Verfahren zu einem eigenständigen Fachgebiet der Statistik beziehungsweise der Informatik entwickelt [vgl. Bec+16, 28 f.].

Der Begriff der Datenvisualisierung umschreibt die visuelle Repräsentation und Präsentation von Daten, um das Verständnis zu verbessern [vgl. Kir19, 15 ff.]. Er wird in Teilen der Literatur als Oberbegriff für „*Information visualization*“ und „*Scientific Visualization*“ verwendet [vgl. Few09, S. 11].

Datenvisualisierungen haben das Ziel, die Analyse, Exploration und Entdeckung der Daten zu ermöglichen. Sie sollen das Verständnis der dargestellten Daten erleichtern und sind anders als Infographiken¹ nicht primär dafür geschaffen, Geschichten über die Informationen zu erzählen [vgl. Kir19, 20 ff.]. Sie werden vielmehr als Werkzeuge verstanden, die es ermöglichen sollen, Entscheidungen aus den visualisierten Daten zu ziehen [vgl. Cai16, S. 31].

In der Fachliteratur finden sich verschiedene Eigenschaften von Datenvisualisierungen. Cairo führt fünf Eigenschaften auf: *truthful*, *functional*, *beautiful*, *insightful* und *enlightening*. Datenvisualisierungen basieren auf gründlicher und ernsthafter Forschung (*truthful*). Sie sind funktional, das heißt sie bemühen sich die Daten genau darzustellen (*functional*). Indem Datenvisualisierungen schwer entdeckbare Beweise offenbaren, sind sie aufschlussreich (*insightful*). Darüber hinaus sollen sie für die Zielgruppe attraktiv sein (*beautiful*). Zudem sind Datenvisualisierungen aufklärend (*enlightening*), da sie Veränderungen im Denken anstoßen können [vgl. Cai16, S. 45].

Die visuelle Repräsentation der Daten erfolgt unter Verwendung graphischer Markierungen (*marks*) wie Punkt-, Linien und Balkensymbolen. Die Eigenschaften dieser Markierungen wie ihre Form, Größe oder Farbe kodieren die darunter liegenden Datenwerte. Die so kodierten Datenwerte werden dann in Diagrammen dargestellt [vgl. Kir19, 135 ff.].

Die visuelle Datenrepräsentation wird von verschiedenen Faktoren beeinflusst. Grundsätzlich ist zu überlegen, ob die Daten in Diagrammen oder Tabellen repräsentiert werden sollen. Daran anschließend ist die Frage zu klären, welche Art der Beziehung zwischen den Daten gezeigt werden soll. Für die Auswahl der Diagramme ist es wichtig zu bestimmen,

¹ Infographiken haben die Aufgabe Nachrichten zu kommunizieren. Sie bestehen aus einer Mischung von Diagrammen, Karten, Illustrationen und Text. Klarheit und Tiefe der Darstellungen sind dabei wichtig [vgl. Cai16, S. 31]. Sie werden auch als „*Explanation Graphics*“ bezeichnet und bestimmen sich dadurch, dass sie Geschehen und Ereignisse graphisch darstellen. Historisch sind Infographiken mit dem Medium der Printzeitungen und Printzeitschriften verbunden [vgl. Kir19, S. 27].

ob ein Kategorienvergleich, eine Zeitreihe, eine Rangfolge, eine relative Häufigkeit oder eine Korrelation dargestellt werden soll [vgl. Few12, S. 137].

Für die Auswahl der richtigen Datenvisualisierung sind des Weiteren die unterschiedlichen Datentypen von großer Relevanz. Datentypen „...define the nature of the values held under each variable and about each item in your dataset.“ [Kir19, S. 99] Eine Variable (Merkmal) kann qualitativ (kategorial) oder quantitativ (metrisch) sein. Die [Abbildung 2.1](#) zeigt die statistischen Datentypen nach der Nominal-, Ordinal-, Intervall-, Ratio-Systematik (NOIR) [vgl. BS10, 12 ff.] mit möglichen Aussagegehalten und Beispielen.

2



Abbildung 2.1: Statistische Datentypen mit Aussagegehalten und Beispielen

Unterschieden werden die Merkmale ferner nach diskret und stetig. Ein diskretes Merkmal kann auf der Basis der natürlichen Zahlen abzählbar viele Merkmalsausprägungen annehmen. So ist zum Beispiel die Größe des Medienbestandes einer Bibliothek ein diskretes Merkmal, da es keine halben oder viertel Medien gibt. Im Gegensatz dazu können die Merkmalsausprägungen eines stetigen Merkmals jeden beliebigen Wert annehmen. So ist

² Manchmal ist auch nur die Unterscheidung zwischen nominalen, ordinalen und metrischen Merkmalen in der wissenschaftlichen Literatur anzutreffen [vgl. Cle11, S. 20]. Unter Berücksichtigung der großen Vielfalt (variety) der Daten, schlägt Kirk eine Erweiterung der NOIR-Systematik um einen textuellen Datentyp vor [vgl. Kir19, S. 100].

zum Beispiel die Raumtemperatur ein stetiges Merkmal. Qualitative Merkmale können nur diskret sein, während quantitative Merkmale sowohl diskret als auch stetig sein können [vgl. [Kir19](#), 102 f.]. Für kategoriale Merkmale eignen sich Balken- und Kreisdiagramme eher als Liniendiagramme, da kategoriale Daten nur diskret sein können.

Der Einsatz von Datenvisualisierungen wird außerdem bestimmt von der Größe der Datenmenge. Die Darstellung einer großen Datenmenge kann zum Beispiel durch die Wahl eines bestimmten Diagrammes überladen wirken. So stößt die Darstellung einer Datenmenge mit vielen Kategorien durch Balken- und Kreisdiagramme in der Übersichtlichkeit an Grenzen und kann den Wert der zu erzielenden Aussage verwischen [Vgl. [Few12](#), 5 ff.].

Daten können softwareseitig mit Tabellenkalkulationsprogrammen visualisiert werden. In Data-Science-Projekten können außerdem verschiedene Frameworks zum Einsatz kommen. Populär sind die Bibliotheken Matplotlib, Seaborn, Plotly für die Programmiersprache Python. Für die Programmiersprache R gibt es ebenfalls vielfältige Möglichkeiten der Visualisierung mit ggplot. Einen Überblick zeigt die Webseite „The Chart maker“ [[WK20](#)]. Eine mächtige und verbreitete JavaScript-Bibliothek zur Datenvisualisierung ist D3.js.

Daten entstehen in wissenschaftlichen Experimenten, aus statistischen Erhebungen wie dem Census, auf Kassenzetteln, in Smartphones, in Unternehmen oder in Einrichtungen wie Bibliotheken. In Unternehmen werden Datenvisualisierungen eingesetzt, um auf der Managementebene die operative Planung und Entscheidungsfindung zu unterstützen. Dabei ist Datenvisualisierung ein Bestandteil eines ganzheitlichen Prozess, der unter dem Begriff Business Intelligence (BI) fungiert.

2.3 BUSINESS-INTELLIGENCE-SYSTEME

Business Intelligence (BI)³ bezeichnet allgemein Konzepte und Methoden zur Entscheidungsfindung, die auf erfassten Informationen beruhen. *BI-Systeme* gehören zu den Management-Support-Systemen, die seit den 1960er Jahren in Unternehmen im Einsatz sind [vgl. Gro20, S. 83].

Populär wurde der Begriff *BI* in den 1990er Jahren. Die Verbreitung des Begriffs fiel zusammen mit der Entwicklung einheitlich strukturierter und dauerhaft verfügbarer Datenbanken in Unternehmen, sogenannten Data Warehouses (DWH). Grund für diese Entwicklung waren die immer neuen Informationsbedürfnisse auf Managementebene, die schnell befriedigt werden sollten. Die Einrichtung dieser Datenbanken zielt ab auf eine umfassende Informationsversorgung durch Verdichtung der Unternehmensdaten [vgl. AM17, 268 ff.].

BI beschreibt einen integrierten, unternehmensspezifischen, IT-basierten Gesamtansatz zur Unterstützung betrieblicher Entscheidungen [vgl. AM17, S. 270]. Ein Hauptmerkmal der *BI* ist nach Linden die Entscheidungsunterstützung der Managementebene [vgl. Lin16, S. 111]. Die Abgrenzung zu operativen Anwendungssystemen wie Online Transaction Processing (OLTP)-Systemen ist laut Abts und Mülder ein weiteres wichtiges Merkmal der *BI-Systeme* [vgl. AM17, S. 267].

Die wesentlichen Schichten eines *BI-Systems* umfassen die Bereiche der internen und externen Datenlieferanten, den Komplex der Datenintegration und -aufbereitung, das Gebiet der Datenspeicherung - und bereitstellung im *Data Warehouse* und den Zweig der Datenanalyse und -präsentation [vgl. Lin16, 126 ff. vgl. KBM10, S. 8].

Die Schichten der *BI-Systeme* können in einem Referenzarchitekturmodell abgebildet werden. Die Referenzarchitektur kann auch die Datenflüsse zwischen den verteilten Systeme-

³ Synonym wird auch manchmal der Begriff Analytische Informationssysteme oder Business-Intelligence-Systeme gebraucht

Einheiten und Komponenten erfassen [vgl. [Lin16](#), 126 ff.]. Eine schematische Darstellung in Form eines solchen Referenzarchitekturmodell zeigt die [Abbildung 2.2](#).



Abbildung 2.2: Schichten eines Business-Intelligence-Systems

Von den internen und externen Datenlieferanten werden die Daten im Bereich der Datenintegration und -aufbereitung mithilfe von *ETL-Prozessen* bearbeitet. In einem ersten Schritt werden die Daten aus den *OLTP-Systemen* oder aus anderen Datenquellen extrahiert. Der anschließende Transformationsprozess wandelt die Daten in ein homogenes Format um. Dabei handelt es sich um einen vierstufigen Prozess. Die Daten werden mit Hilfe zum Teil automatischer Verfahren bereinigt, harmonisiert, verdichtet (aggregiert) und angereichert. Bereinigt werden die Daten unter anderem von syntaktischen und semantischen Mängeln. Unterschiedliche Codierungen der Daten werden durch die Harmonisierung beseitigt. Des Weiteren werden die Daten verdichtet, das heißt es werden Summationen von den Daten auf verschiedenen Ebenen durchgeführt und gespeichert.

Die Berechnung und Speicherung wichtiger Kennzahlen geschieht durch das Verfahren der Anreicherung [vgl. Gro20, S. 86; vgl. AM17, 277 f.]. Schließlich werden die Daten in einem Ladeprozess in das *Data Warehouse* geladen [vgl. Lin16, 129 ff.].

Die Datenspeicherung und -bereitstellung erfolgt im *DWH*. Dieser Bereich ist nach Linden zentral für die *BI-Referenzarchitektur* [vgl. Lin16, S. 135]. Ein *Data Warehouse* ist ein logisch zentralisiertes Datenhaltungssystem. Dieses ist physisch von den operativen Anwendungssystemen getrennt und stellt eine harmonisierte Datenbasis für betriebswirtschaftliche Analysen bereit [vgl. MB00, S. 135].⁴ Ein *Data Warehouse* zeichnet sich durch die vier Merkmale themenorientiert, zeitorientiert, integriert und nicht-volatil aus [vgl. Inm05, 29 f. vgl. AM17, 271 f. vgl. Lin16, 136 f.]. Die Speicherung der Daten erfolgt nach Themenschwerpunkten und orientiert sich am Informationsbedarf des Unternehmens. Die zeitorientierte Speicherung der Daten ermöglicht Zeitreihenanalysen auf historischen Daten. Für die Schaffung einer homogenen Datenbasis integriert das *DWH* die Daten aus heterogenen Datenquellen [vgl. Lin16, S. 136].

Die Datenintegration in das *Data Warehouse* kann auf Grundlage multidimensionaler Datenmodelle wie sogenannten Datenwürfeln erfolgen. Multidimensionale Datenwürfel können n-Dimensionen haben. Sie bestehen mit Fakten und Dimensionen aus zwei Elementen. Fakten sind Kennzahlen, die in den Zellen des multidimensionalen Würfels enthalten sind. Dimensionen sind Entitäten, die um einen Fakt angeordnet sind. Die Dimensionen spannen die Kanten des multidimensionalen Datenwürfels auf. Die Betrachtung der Kennzahlen ist so in einem multidimensionalen Datenmodell anhand verschiedener Dimensionen möglich [vgl. Far11, 13 ff., 21 f. vgl. KBM10, 66 f.].

Ein *Data Warehouse* bietet des Weiteren eine nicht-volatile Speicherung der Daten an. Damit ist sichergestellt, dass auf diesen Daten längerfristige Analysen durchgeführt werden können [vgl. Lin16, S. 136]. In die Konzeptionierung eines *Data Warehouses* sollten deswegen Archivierungskonzepte und zudem Überlegungen zu den Aktualisierungszyklen

⁴ Alternativen zum *Data Warehouse* wären verschiedene *Data Marts*, die kleinere Datenspeichereinheiten darstellen und sich inhaltlich an späteren Abfrage- und Auswertungszwecken orientieren.

der Daten für das *DWH* miteinfließen. Die Archivierungskonzepte sorgen dafür, dass veraltete Datenbestände gesichert und komprimiert werden. Die Aktualisierungszyklen legen fest, in welcher periodischen Abfolge die Daten aktualisiert werden, entweder zu Zeitpunkten der Änderungshäufigkeit der Daten im operativen Anwendungssystem, in periodischen Zeitabständen oder aber auch in Echtzeit [vgl. [Lin16](#), S. 137].

Bei der Umsetzung mit Datenbanktechnologien besteht die Möglichkeit, das multidimensionale Modell in einer relationalen Datenbank umzusetzen. Realisiert werden kann dies mit einer logischen Datenmodellierung durch ein Star- oder Snowflake-Schema [vgl. [Lin16](#), 177 f.].

Die letzte Schicht eines *BI-Systems* umfasst die Datenanalyse und -präsentation. Die Datenanalyse kann verschiedene Auswertungskonzepte aufweisen. Die Auswertungskonzepte bieten spezifische Funktionen für Analysen an. Im Rahmen herkömmlicher *BI-Systeme* werden Online Analytical Processing (OLAP)-Verfahren angewendet. *OLAP* ist eine Anfragetechnik für die Analyse multidimensionaler oder relationaler Daten. Diese Technik erlaubt es, Daten mit *OLAP*-Funktionen wie Drill-Down, Roll-up, Slice oder Dice auf verschiedenen Stufen mit unterschiedlichen Sichtweisen darzustellen. So ist eine Verfeinerung der Analyseergebnissen (Drill-Down) und deren Aggregation möglich. Darüber hinaus können Teilmengen durch Slicing und Dicing gebildet werden [vgl. [AM17](#), 283 f.]. Durch *OLAP*-Anwendungen können die multidimensionalen Datenstrukturen interaktiv ausgewertet werden.

Ein anderes Verfahren ist Data Mining. Data Mining benutzt verschiedene statistische und mathematische Verfahren, um Muster und Trends in vor allem großen Datenmengen zu entdecken. Klassische Data-Mining-Aufgaben sind Ausreißer-Erkennung, Klassifikation, sowie die Cluster-, Assoziations- und Regressionsanalysen [vgl. [HKP12](#), 15 ff.]. Data Mining wird entweder synonym gesetzt mit dem Prozess der Knowledge Discovery in Databases (KDD) oder als Teilphase dieses Prozesses beschrieben [vgl. [HKP12](#), S. 6; vgl.

[Lin16](#), 142 f.]. Data-Mining-Verfahren können als zusätzliche Auswertungsverfahren zu *OLAP*-Verfahren hinzutreten.

Die Datenpräsentation umfasst die strukturierte und visuelle Darstellung der zuvor angewendeten Analyseverfahren. Die Präsentation der Daten kann durch Reporting oder Dashboards erfolgen. Über Reporting-Tools können Standardberichte generiert werden. Diese statischen Berichte können nicht verändert werden. Für die Darstellung der Daten in diesen Berichten werden Tabellen, Listen und Diagramme verwendet. Der Bericht kann bereitgestellt werden in PDF-Format [vgl. [Ban16](#), S. 114].

Im Gegensatz dazu ermöglichen Dashboards einen interaktiven Zugang über eine Benutzungsoberfläche zu den relevanten Informationen. Nach Few ist ein Dashboard ein „... visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.“ [[Few06](#), S. 26] Dashboards werden im unternehmerischen Rahmen auch Performance Dashboards genannt [vgl. [Lin16](#), S. 154]. Diese werden verschieden klassifiziert. Sie lassen sich nach Reichweite und Zweck einteilen. Es gibt operative, taktische (analytische) oder strategische Dashboards.

Operative Dashboards überwachen und kontrollieren gegenwärtige Geschäftsprozesse. Durch die hohe Aktualisierungsfrequenz der Daten kann schnell in die betrieblichen Prozesse steuernd eingegriffen werden [vgl. [Eck11](#), 11 f. vgl. [Few06](#), 30 f.].

Taktische oder analytische Dashboards konzentrieren sich auf verschiedene Bereiche des Unternehmens und bieten Möglichkeiten einer anforderungsgerechten und lösungsorientierten Analyse durch eine Bewertung der Daten auf mehreren Detailebenen.

Strategische Dashboards repräsentieren hochaggregierte Kennzahlen, die langfristige Ziele und deren Erreichungsgrade visualisieren. Sie werden zur Kommunikation und Kollaboration auf oberster Managementebene genutzt [vgl. [Lin16](#), 155 f.].

Die Grenzen zwischen den Dashboards bezüglich Reichweite und Zweck sind aber nicht trennscharf. So kann es Überlappungen zwischen strategischen, taktischen und operativen Dashboards geben [vgl. [Eck11](#), S. 121].

Neuere Entwicklungen in den *BI-Systemen* führen weg von der strikten Trennung zwischen OLTP und OLAP und zur Auflösung der *DWH*. Anstelle derer treten Data Lakes. Data Lakes speichern im Gegensatz zu den Data Warehouses die Rohdaten in strukturierter oder unstrukturierter Form. Mitunter wird dabei auf die aufwändigen *ETL-Prozesse* verzichtet [vgl. [Gro20](#), S. 86], und die Daten werden zum Analysezeitpunkt bearbeitet.

3 AUSGANGSSITUATION

Im folgenden Kapitel wird die wissenschaftliche Spezialbibliothek des *Max-Planck-Institutes für empirische Ästhetik* porträtiert, um die Ausgangslage zu umreißen. Anschließend werden die bibliothekarischen Informationsdienstleistungen der Bibliothek skizziert und der Frage nachgegangen, welche statistischen Daten aggregiert und ausgewertet wurden. Diese Daten stellen die Basis für die spätere Konzeption und Entwicklung eines datengetriebenen Unterstützungssystems dar.

3.1 BIBLIOTHEK

3.1.1 ALLGEMEINES

Die Spezialbibliothek wurde im Zuge der Gründung des *MPI EA* in Frankfurt im Jahr 2013 gegründet. Die Aufgabe des Institutes ist die interdisziplinäre Erforschung empirischer Fragestellungen der Ästhetik. Das Institut besteht derzeit aus drei Abteilungen *Sprache und Literatur*, *Musik* und *Neurowissenschaften* sowie einigen Forschungsgruppen.

Die Bibliothek ist eine Serviceeinrichtung des Institutes und dient mit ihren Informationsdienstleistungen der Forschung. Zentral ist dabei die Informationsversorgung der Forschenden. Die benötigten Informationen sind Bücher, Zeitschriften, Zeitschriftenartikel sowohl in gedruckter als auch in elektronischer Form. Der Bibliotheksbestand ist somit hybrid. Er besteht sowohl aus gedruckten als auch Online-Medien sowie audiovi-

suellen Materialien. An Bestand umfasst die Bibliothek zirka 11.000 Bücher, 30 laufende Zeitschriften, knapp über 200 audiovisuelle Medien sowie Online-Datenbanken und Online-Zeitschriften.

3.1.2 ORGANISATORISCHE EINBETTUNG

Um alle Informationsbedarfe der Forscher:innen zu befriedigen, wird die Bibliothek in ihren Aufgaben von der *max-planck-digital-library* unterstützt. Deren Portfolio umfasst vorrangig die zentrale Lizenzierung von relevanten elektronischen Informationsressourcen, die Bereitstellung von Softwarelösungen, das Betreiben des Publikationsrepositoriums *Pu-Re.MPG* der *Max-Planck-Gesellschaft* (MPG) sowie das Vorantreiben von Open-Access-Initiativen.

Darüber hinaus ist die Spezialbibliothek Teil des *HeBis-Verbundes*. Seit Ende 2014 finden die Geschäftsprozesse der Katalogisierung und der Erwerbung im *Zentralsystem* (CBS) und im Lokalsystem *Lokalsystem* (LBS) vom *OCLC* statt. Im *Online-Katalog* (OPAC) befinden sich Bücher, ausgewählte E-Books und Zeitschriften (Print und Online) der Institutsbibliothek. Lokal lizenzierte Datenbanken finden sich dagegen nicht im Katalog. Das *LBS* wird gehostet und betreut vom Lokalsystem-Team Frankfurt. Als Service-Leistungen werden der Bibliothek besondere Funktionalitäten für das *Zentralsystem* und Statistiken aus dem *LBS* von dem Lokalsystem-Team bereitgestellt.

3.1.3 INFORMATIONSDIENSTLEISTUNGEN

Das Bibliotheks-Team des *MPI EA* ist verantwortlich für den Ablauf und Organisation der bibliothekarischen Informationsdienstleistungen. Eine Übersicht der Informationsdienstleistungen, aufgeschlüsselt nach den Basisfunktionen einer Bibliothek [Rös+19, S. 204 f.], zeigt [Tabelle 3.1](#). Die zentralen Informationsdienstleistungen der Spezialbibliothek bestehen aus der Sammeltätigkeit und dem Benutzungsservice. Seit der Institutsgründung wird neben dem nutzungsorientiertem Bestandaufbau ebenfalls eine planmäßige

Bestandsentwicklung betrieben. Das Erwerbungsprofil der Bibliothek leitet sich aus dem Forschungsauftrag des Institutes ab und umfasst dementsprechend die Erwerbung von Informationsressourcen, die sich den theoretischen und empirischen Fragestellungen der Ästhetik widmen.

Die Dienstleistungsbereiche der Benutzung sind zuständig für die Organisation der Fern- und Ortsleihe von Informationsressourcen, die nicht in das Erwerbungsprofil der Spezialbibliothek fallen. Ferner sind diese für die Informationsbeschaffung sowohl über Dokumentenlieferdienste als auch für die Akquise von einzelnen Zeitschriftenaufsätzen zuständig.

Basisfunktion	Beschreibung
Benutzung	Ausleihe, Lesesaalnutzung, Organisation der Lieferdienste (Fern und Ortsleihe, Dokumentenlieferdienste)
Management techn. Infrastruktur	<i>PuRe.MPG</i> , Medien-Datenbank
Ordnen	Aufstellungssystematik (RVK)
Sammeln und Erschließen	geplanter Bestandsaufbau, Integrierter Geschäftsgang Medienerwerbung und Medienschließung, besondere Materialien
Vermitteln	Literaturrecherche, Nutzung elektronischer Ressourcen, Urheberrecht und Publikationsberatung

Tabelle 3.1: Informationsdienstleistungen nach Basisfunktionen der Spezialbibliothek

Weitere Informationsdienstleistungen sind die Betreuung des Publikationsrepositoriums *PuRe.MPG* des Institutes, spezielle Beratungsdienstleistungen zum Urheberrecht und zum Publishing sowie klassische Auskunft- und Informationsdienste. Seit Beginn 2016 geschieht die Ausleihe der Medien über ein Selbstverbuchungssystem.

3.1.4 EVALUATION DER INFORMATIONSDIENSTLEISTUNGEN

Zu fast jeder Informationsdienstleistung der Spezialbibliothek werden quantitative Daten elektronisch generiert. [Tabelle 3.2](#) zeigt Daten, die bereits jetzt in der ein oder anderen Form aggregiert und ausgewertet werden. Die Tabelle stellt nach den Evaluationstypen

3 Ausgangssituation

dar, ab wann und wie regelmäßig die Statistiken erfasst werden. Ferner bietet sie einen Überblick darüber, in welchem Format die Daten vorliegen, über die Quellen aus der die Daten stammen und ob die Daten bereits systematisch ausgewertet und/oder visualisiert werden.

Evaluationstyp	Basisfunktion	Daten	Zeitraum	Turnus	Quelle	Format	Syst. Auswertung	Visualisierungen
Nutzungsbezogen	Benutzung	Ausleihzahlen Bibliotheksbestand	2016-	monatlich	LBS	Mail, xlsx	nein	-
Nutzungsbezogen	Benutzung	Ausleihzahlen Lieferdienste	2015-	monatlich	intern	xlsx	ja	teilweise, Liniendiagramm
Nutzungsbezogen	Benutzung	Besonders nachgefragte Medien (OPAC)	2017-	monatlich	LBS	Mail, txt	nein	-
Nutzungsbezogen	Sammeln	COP 5-Statistiken elektr. Ressourcen	2013-	unbekannt	mpdl	csv, tsv, txt	nein	-
Nutzungsbezogen	Benutzung	Lesesaalnutzung	2015-	wöchentlich	intern	xlsx	nein	-
Sammlungsbezogen	Sammeln	Budget nach Kostenstellen	2018-	monatlich	LBS	Mail, txt	ja	-
Sammlungsbezogen	Sammeln	Umsatz nach Lieferanten	2018-	monatlich	LBS	Mail, txt	ja	Balken- und Kreisdiagramm
Sammlungsbezogen	Sammeln	Größe und Art des Bestandes	2014-	jährlich	LBS, intern	csv	nein	-
Sammlungsbezogen	Sammeln	Neuerwerbungslisten	2014-	monatlich	LBS, intern	tsv	nein	-

Tabelle 3.2: Liste der Dienstleistungsbereiche zu denen statistische Daten erhoben werden

Intern erfasst die Bibliothek monatlich die Daten der Ausleihe über die Lieferdienste. Unterschieden wird in der Erfassung nach Medientypen, nach Ausleihort und Ausleihart. Zudem werden wöchentlich Statistiken zur Nutzungshäufigkeit des Lesesaals geführt. Jährlich wird die Bestandsgröße nach Medientyp für die Buchhaltung ermittelt.

Die Ausleihzahlen des Bibliotheksbestandes werden bei Bedarf durch das Lokalsystem-Team ermittelt und an die Bibliothek geschickt. Diese liegen kumulativ nach Ausleihanzahl des einzelnen Titels oder nach Jahr und der Identifikationsnummer des Titeldatensatzes im CBS vor. Ebenfalls stehen die Ausleihzahlen als Rohdaten, in denen jede Titelausleihe aufgeführt wird, zur Verfügung.

Anzahl von cum_loans		Spaltenbeschriftungen					
Zeilenbeschriftungen		2016	2017	2018	2019	2020 (Leer)	Gesamtergebnis
5	74290844			1			1
6	74294621			1	1		2
7	74294826					1	1
8	74294926			1	1	1	3
9	74294971			1	1		2
10	74295032				1	1	2
11	74295067			1	1		2
12	74295100		1	1	1	1	4
13	74295652		1	1	1	1	4
14	74295755		1	1		1	4
15	74298047		1	1			2

Abbildung 3.1: Anzahl der Titelausleihen nach Identifikationsnummer und Jahr

Monatlich bekommt die Bibliothek kumulative Budget- und Umsatzübersichten der Kostenstellen und der Lieferanten zugeschickt. Die Budgetübersicht gibt den monatlichen Stand der Budgets (Bindungen und Ausgaben) der einzelnen Kostenstellen zwischen den ACQ-Jahresübergängen aus.⁵ Die Kostenstellen bilden die einzelnen Abteilungen und zum Teil die Forschungsgruppen des Institutes ab.

Die Umsatzübersicht summiert die monatlichen Ausgaben nach Lieferanten bezogen auf das Kalenderjahr. Kriterium ist hierbei das Erfassungsdatum der Rechnung.

Bearbeitet werden nur die Umsatzübersichten der Lieferanten, um die Umsatzverteilung nach Lieferanten zu steuern. [Abbildung 3.2](#) zeigt ein Beispiel⁶ vom November 2020 für die Budget- und Umsatzübersichten des Monats Oktober 2020, die vom Lokalsystem-Team als email zur Verfügung gestellt werden.

Budgetuebersicht Bibliothek MPI fuer Empirische Aesthetik						Umsatzuebersicht Bibliothek MPI fuer Empirische Aesthetik	
Stand: Nov 1 2020 6:16AM						Stand: Nov 1 2020 6:16AM	
Budgetuebersicht alle Budgets						Lieferant Umsatz (EUR)	
S	Bezeichnung	Ansatz	Bindungen	Ausg. ges.	Bestellvol.		
bet m 20	Betriebsrat	0.00	0.00	0.00	0.00	Antiquariat allgemein	0.00
bib m 20	Bibliothek	0.00	0.00	0.00	0.00	Autorenbuchhandlung Marx & Co GmbH	0.00
bib z 20	Bibliothek	0.00	0.00	0.00	0.00	COMPUTEC MEDIA GmbH	0.00
fam m 20	Fachgeschichte Musik	0.00	0.00	0.00	0.00	Deutscher Hochschulverband	0.00
hmm z 20	Histories of Music Mind and Body	0.00	0.00	0.00	0.00	EBSCO Information Services GmbH	0.00
inf m 20	IT	0.00	0.00	0.00	0.00	EBSCO Information Services GmbH Buch	0.00
lit m 20	Literatur und Sprache	0.00	0.00	0.00	0.00	Erasmus Boekhandel bv	0.00
mus m 20	Musik	0.00	0.00	0.00	0.00	Ersatzlieferant	0.00
neu m 20	Neurowissenschaft	0.00	0.00	0.00	0.00	Felix Meiner Verlag GmbH	0.00
vew l 20	Verwaltung	0.00	0.00	0.00	0.00	Heinemann Verlag Shop	0.00
vew m 20	Verwaltung	0.00	0.00	0.00	0.00	Karl-Marx-Buchhandlung GmbH	0.00
wid m 20	Wissenschaftliche Dienste	0.00	0.00	0.00	0.00	Massmann Internationale Buchhandlung GmbH	0.00
						Missing Link Versandbuchhandlung eG	0.00
						Presse Verlagsgesellschaft fr Zeitschriften und neue Medien mbH	0.00
						Schweitzer Fachinformationen	0.00
						Springer	0.00
						Texte zur Kunst GmbH & Co. KG	0.00
						Verlagsgruppe Hthig Jehle Rehm	0.00
						ebooks	0.00
Summe		0.00	0.00	0.00	0.00	Umsatz (EUR)	0.00

Abbildung 3.2: Monatliche Budget- und Umsatzübersicht

Die Ausgaben für die lokal lizenzierten Datenbanken fehlen in der Aufstellung der Ausgaben und werden in einer Tabelle extra geführt.

⁵ Der ACQ-Jahresübergang variiert und der Zeitpunkt wird von der Bibliothek vereinbart mit dem Lokalsystem-Team. Der Jahresübergang geschieht meistens in den ersten beiden Januarwochen des Jahres.

⁶ Die Zahlen wurden zurückgesetzt.

3 Ausgangssituation

Die *Counter 5-Statistiken* (COP 5) der Verlage werden auf einen internen Portal von der *mpdl* dem Institut zur Verfügung gestellt. Diese Statistiken verzeichnen den Zugriff innerhalb der IP-Adressbereiche des Institutes auf die elektronische Ressourcen, die konsortial durch die MPG lizenziert wurden. Darunter fallen ebooks der Verlage *Springer*, *Wiley* oder *De Gruyter*. Bisher wurden diese von der Bibliothek noch nicht gesichtet und ausgewertet.

Eine proaktive und systematische Auswertung der Entwicklung der Bestandsgröße, der Ausleihzahlen und der *COP 5*-Statistiken findet nicht oder nur unzureichend statt. Auch wird das Potential wie in [Abschnitt 2.1](#) beschrieben hinsichtlich der Budgetplanung nicht ausgeschöpft.

4 KONZEPTION EINER LÖSUNG

Ausgehend von der Analyse der Ausgangssituation werden im Folgenden Anforderungen für das zu entwickelnde System herausgearbeitet. Alle Anforderungen dienen als Grundlage für die Entwicklung des Proof-of-Concepts. Schließlich werden 7 Anwendungsfälle für das System im [Unterabschnitt 4.1.5](#) beschrieben, die es zu erfüllen hat. Anhand der Anforderungen und der Anwendungsfälle erfolgt im [Kapitel 5](#) die Bewertung des Proof-of-Concepts.

4.1 ANFORDERUNGSANALYSE

Die Vision und die Ziele des Systems werden zunächst kurz formuliert. Im Anschluss daran, werden die Rahmenbedingungen erarbeitet. Danach werden sowohl die funktionalen Anforderungen als auch die nicht-funktionalen Anforderungen festgelegt.

Die Anforderungen werden nach dem *MoSCoW-Prinzip* priorisiert. Hierbei wird unterschieden in Muss-Anforderungen (M - must), in Soll-Anforderungen (S - should), in Kann-Anforderungen (C - could) und in Anforderungen, die im Zuge der Implementierung des Proof-of-Concepts (noch) nicht umgesetzt werden (W - would / won't). Soweit nötig wurden für die Rahmenbedingungen und für die nicht-funktionalen Anforderungen Messbarkeitskriterien formuliert.

4.1.1 VISION UND ZIELE

Die Bibliothek des *MPI EA* soll durch das System in die Lage versetzt werden, ihr Budget effizient und bedarfsgerecht zu planen. Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen können sich relevante *Key Performance Indicators* durch das System mit Datenvisualisierungen anzeigen lassen. Relevante *KPI* sind die Budget- und Umsatzübersichten, die Ausleihzahlen, das Bestandswachstum und die Neuerwerbungen. Ausgewählte *KPI* werden an die Institutsleitung als Standardreport verteilt.

4.1.2 RAHMENBEDINGUNGEN

Die Rahmenbedingungen legen die organisatorischen Anforderungen für das zu entwickelnde System fest. Darunter fallen die Anwendungsbereiche, die unterschiedlichen Zielgruppen und die technischen Anforderungen.

ID	Beschreibung	Messbarkeit	Priorisierung
R1	Zielgruppen für das System sind die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen.	-	M
R2	Die Institutsleitung ist die Zielgruppe für den Standardbericht, der aus dem System erzeugt wird.	-	M
R3	Das System wird als Desktop-Anwendung in einer Büroumgebung eingesetzt.	Das System wird nicht optimiert für mobile Endgeräte wie Smartphones oder Tablets.	M
R4	Das System wird bedarfsorientiert gestartet und beendet.	-	M
R5	Der Betrieb des Systems läuft unbeaufsichtigt.	-	M
R6	Die Entwicklungsumgebung kann identisch mit der Produktivumgebung sein.	Vergleich der Entwicklungsumgebung mit der Produktivumgebung.	C
R7	Die Softwareanforderungen sind für das System dokumentiert.	Überprüfen, ob entsprechendes Dokument vorhanden ist.	M
R8	Die Hardwareanforderungen sind für das System dokumentiert.	Überprüfen, ob entsprechendes Dokument vorhanden ist.	M
R9	Das System läuft in einem geschützten Bereich im internen Netzwerk des Institutes, der nur für das Bibliothekspersonal einsehbar ist.	Testen des Zugriffs von außerhalb des instituts-eigenen IP-Adressbereich und von bibliotheks-fremden Mitarbeiter:innen des Instituts.	W
R10	Das System wird auch von anderen Bibliotheken mit einer ähnlichen Datenlage eingesetzt.	Testen des Systems durch eine andere Bibliothek	W

Tabelle 4.1: Rahmenbedingungen

4.1.3 FUNKTIONALE ANFORDERUNGEN

Im Folgenden werden die funktionalen Anforderungen an das System formuliert. Das System strukturiert sich in die Bereiche des *ETL-Prozesses* und der Datenspeicherung, der Datenanalyse, der Datenpräsentation und des Standardberichtes.

Die [Tabelle 4.2](#) beschreibt die Anforderungen an das System bezüglich des *ETL-Prozesses* und der Datenspeicherung.

ID	Beschreibung	Priorisierung
F1	Das System importiert automatisch die Daten von einem lokalem Verzeichnis.	M
F2	Das System bereinigt automatisch die Daten von syntaktischen Fehlern und vollzieht Formatanpassungen (einheitliches Datumsformat, einheitliches Zeichenformat, einheitliches Tabellenformat).	M
F3	Das System harmonisiert die Daten automatisch (Erkennen und Harmonisierung von unterschiedlichen Codierungen der Datenwerte, Erkennen und Zusammenführung von Synonymen).	M
F4	Das System ergänzt die Daten mit zusätzlichen Daten aus anderen Datenquellen (z.B. RVK-Systematik).	M
F5	Das System speichert automatisch die Daten an einem definierten Speicherort.	M
F6	Das System speichert automatisch redundant die Daten für ein Backup an einem definierten Ort.	S
F7	Das System löscht nach dem Importprozess automatisch die Daten aus dem lokalen Verzeichnis.	C
F8	Das System dokumentiert zur Fehlererkennung den <i>ETL</i> - und den Speicherprozess in einer log-Datei.	W

Tabelle 4.2: Funktionale Anforderungen - ETL-Prozess und Datenspeicherung

[Tabelle 4.3](#) legt die Anforderungen für die Datenanalyse an das System fest.

ID	Beschreibung	Priorisierung
F10	Das System bietet eine Graphische Benutzeroberfläche (GUI) an.	M
F11	Das System analysiert die Daten automatisch mit deskriptiven und explorativen Methoden der Statistik nach Kriterien, die in den Anwendungsfällen formuliert sind.	M
F12	Das System filtert die Daten nach bestimmten Kriterien, die in den Anwendungsfällen formuliert sind.	M
F13	Das System bietet die Filterung der Daten über die <i>GUI</i> auf Basis von Filtern, die von den Benutzer:innen eingegeben werden, an.	M
F14	Das System analysiert die gefilterten Daten automatisch mit deskriptiven und explorativen Methoden der Statistik nach Kriterien, die in den Anwendungsfällen formuliert sind.	M

Tabelle 4.3: Funktionale Anforderungen - Datenanalyse

4 Konzeption einer Lösung

Die Anforderungen für die Datenpräsentation und die Erstellung des Standardberichtes sind in [Tabelle 4.4](#) aufgelistet.

ID	Beschreibung	Priorisierung
F15	Das System bietet für die betreffenden Daten Datenvisualisierungen an.	M
F16	Die angebotenen Datenvisualisierungen sind hauptsächlich Linien- und Balkendiagramme.	M
F17	Das System bietet für die betreffenden Daten eine Auswahl an diesen Datenvisualisierungen an.	S
F18	Das System erstellt bedarfsorientiert ein PDF-Dokument mit den relevanten <i>KPI</i> .	M
F19	Das System speichert bedarfsorientiert Diagramme der relevanten <i>KPI</i> in einem platzsparenden Bildformat.	C
F20	Das System greift auf Diagrammbilder und Texte für die Generierung des PDF-Dokumentes automatisch zu.	M
F21	Das System öffnet das PDF-Dokument automatisch.	M

Tabelle 4.4: Funktionale Anforderungen - Datenpräsentation und Standardbericht

4.1.4 NICHT-FUNKTIONALE ANFORDERUNGEN

In der [Tabelle 4.5](#) sind die nicht-funktionalen Anforderungen, die Qualitätskriterien an das System beschreiben, aufgelistet.

ID	Beschreibung	Messbarkeit	Priorisierung
NF1	Das System ist portierbar auf eine andere Plattform.	Testen mit unterschiedlichen Systemen (OS, Browser).	M
NF2	Das System ist leicht erlernbar.	Schulungsdauer weniger als 1 Stunde.	M
NF3	Der Zugriff auf das System erfolgt passwortgeschützt.	-	W
NF4	Das System ist modular aufgebaut. Die einzelnen Module sind testbar.	Module sind unabhängig voneinander testbar.	S
NF5	Zur Programmierung des Systems wird freie Software (Programmiersprachen) genutzt.	Überprüfen, ob Software freie Lizenz enthält.	M
NF6	Die eingesetzte Software ist weitverbreitet und geeignet für diese Aufgabe.	Überprüfen der Verbreitung und Eignung der Software.	M
NF7	Das System ist einfach zu warten.	-	S
NF8	Das System ist unter einer Open-Source-Lizenz zu entwickeln.	Vergabe einer freien Lizenz für das zu entwickelnde System.	M
NF9	Die Reaktionszeit des Systems auf Benutzungsanfragen über die GUI beträgt weniger als 2 Sekunden.	Tests mit dem System.	S
NF10	Auf das System kann in vollem Umfang von mehreren Endgeräten gleichzeitig zugegriffen.	Tests mit mehreren Endgeräten, die gleichzeitig auf System zugreifen.	S
NF11	Das Layout des Dashboards ist strukturiert, übersichtlich und selbsterklärend.	Erfassen relevanter Informationen in weniger als 5 Sekunden.	M
NF12	Die Entwicklung des Systems erfolgt mit modernen Technologien.	-	M
NF13	Das Layout des Dashboards ist am Corporate Design des Institutes auszurichten.	-	W
NF14	Die verschiedenen Diagrammtypen werden zielgerichtet eingesetzt.	-	M
NF15	Optische Gestaltungsmerkmale und Farben der Datenvisualisierungen werden zur Verdeutlichung der Information eingesetzt.	-	M
NF16	Optische Effekte oder Animationen der Datenvisualisierungen sind sparsam einzusetzen.	-	M
NF17	Die Darstellung der Informationen ist auf die Zielgruppen zugeschnitten.	-	M
NF18	Die Datenvisualisierung zielt auf die Beantwortung der spezifischen Anwendungsfälle.	-	M

Tabelle 4.5: Nicht-funktionale Anforderungen

4.1.5 ANWENDUNGSFÄLLE

Im folgenden Abschnitt werden die Anwendungsfälle dargestellt, die das System beantworten soll. Für die Entwicklung des Proof-of-Concepts wurde eine Auswahl aus den bibliothekarischen Dienstleistungsbereichen getroffen. Dabei wird auf die in [Tabelle 3.2](#) aufgeführte Evaluationstypen referiert. Jeder Anwendungsfall enthält den Namen des Anwendungsfalls (Titel), den bibliothekarischen Evaluationstyp, die beteiligten Akteur:innen, das zu erreichende Ziel und den Inhalt. Zudem werden Vorbedingungen und die Anforderungen aufgeführt. Das Systemverhalten wird zu den jeweiligen Punkten ebenfalls beschrieben.

Anwendungsfall 1

	Beschreibung	Systemverhalten
Titel	Ausleihzahlen Bibliotheksbestand	-
Evaluationstyp	Nutzungsbezogen	-
Akteur:innen	Bibliotheksführung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige der Ausleihzahlen des Bestandes nach Jahr.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung oder die Bibliotheksmitarbeiter:innen lassen sich auf der <i>GUI</i> die Ausleihzahlen pro Jahr anzeigen.	<p>Das System filtert die betreffenden Datensätze nach Jahr. Das System zeigt diese Datensätze mit Datenvisualisierungen an.</p> <p>Das System zeigt darüberhinaus die ausleihstärksten Titel absteigend nach Anzahl und Jahr an.</p> <p>Das System zeigt darüberhinaus die Verteilung der ausgeliehenen Titel in der <i>RVK</i>-Aufstellungssystematik pro Jahr an.</p> <p>Das System zeigt darüberhinaus die Anzahl der Ausleihen pro Jahr und für den Gesamtzeitraum⁷ im Vergleich zum Bestandswachstum pro Jahr und für den Gesamtzeitraum an.</p> <p>Das System zeigt darüberhinaus die Entwicklung bezüglich der Anzahl der ausgeliehenen Titel im Vergleich zum Gesamtbestand für den Gesamtzeitraum an.</p>
Anforderungen	R1, R4, F2-F4, F10-F17, NF14-NF18	-

Tabelle 4.6: Anwendungsfall 1 - Ausleihzahlen Bibliotheksbestand

⁷ Der Gesamtzeitraum bezieht sich hier und im Folgenden auf den angegebenen Zeitraum in [Tabelle 3.2](#)

Anwendungsfall 2

	Beschreibung	Systemverhalten
Titel	Ausleihzahlen bibliotheksinterne Lieferdienste	-
Evaluationstyp	Nutzungsbezogen	-
Akteur:innen	Bibliotheksleitung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige der Ausleihzahlen der bibliotheksinternen Lieferdienste.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung oder die Bibliotheksmitarbeiter:innen wählen den gewünschten Zeitraum aus und lassen sich die Ausleihzahlen bibliotheksinterner Lieferdienste auf der <i>GUI</i> anzeigen.	<p>Das System filtert die betreffenden Datensätze nach Gesamtzeitraum und Jahr. Das System zeigt diese Datensätze mit Datenvisualisierungen an.</p> <p>Das System zeigt die Nutzung der verschiedenen internen Lieferservices nach Gesamtzeitraum, Jahr und Monat an.</p> <p>Das System zeigt darüberhinaus die verschiedenen internen Lieferservices nach Nutzung der Institutsabteilungen an.</p> <p>Das System zeigt darüberhinaus die Entwicklung in der Nutzung der verschiedenen Lieferservices durch die Abteilungen nach Jahr und im Gesamtzeitraum an.</p>
Anforderungen	R1, R4, F2-F4 F10-F17, NF9, NF14-NF18	-

Tabelle 4.7: Anwendungsfall 2 - Ausleihzahlen bibliotheksinterne Lieferdienste

4 Konzeption einer Lösung

Anwendungsfall 3

	Beschreibung	Systemverhalten
Titel	Lesesaalnutzung	-
Evaluationstyp	Nutzungsbezogen	-
Akteur:innen	Bibliotheksleitung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige der Nutzung des Lesesaals während der Öffnungszeiten.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung oder die Bibliotheksmitarbeiter:innen lassen sich die Nutzung des Lesesaals auf der <i>GUI</i> anzeigen.	Das System filtert die betreffenden Datensätze nach Gesamtzeitraum und Jahr. Das System zeigt diese Datensätze mit Datenvisualisierungen an. Das System zeigt die Nutzung des Lesesaals nach Monat und Jahr, gruppiert in drei Zeitschichten an.
Anforderungen	R1, R4, F2, F3, F10-F17, NF9, NF14-NF18	-

Tabelle 4.8: Anwendungsfall 3 - Lesesaalnutzung

Anwendungsfall 4

	Beschreibung	Systemverhalten
Titel	Neuerwerbungen	-
Evaluationstyp	Sammlungsbezogen	-
Akteur:innen	Bibliotheksführung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige der Anzahl der Neuerwerbungen nach <i>RVK</i> -Systematikstelle pro Monat.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen wählen einen Monat auf der <i>GUI</i> aus.	<p>Das System filtert die betreffenden Datensätze nach Monat. Das System zeigt diese Datensätze mit Datenvisualisierungen an.</p> <p>Das System zeigt die Anzahl der Neuerwerbungen nach Medientyp pro Monat nach <i>RVK</i>-Systematikstelle an.</p> <p>Das System zeigt darüberhinaus die Gesamtzahl der Neuerwerbungen nach Medientyp pro Monat an.</p> <p>Das System zeigt darüberhinaus die Anzahl der monatlichen Neuerwerbungen des laufenden Monats im Vergleich zu den Monaten vergangener Jahre an.</p>
Anforderungen	R1, R4, F2-F4, F10-F17, NF9, NF14-NF18	-

Tabelle 4.9: Anwendungsfall 4 - Neuerwerbungen

4 Konzeption einer Lösung

Anwendungsfall 5

	Beschreibung	Systemverhalten
Titel	Bestandswachstum	-
Evaluationstyp	Sammlungsbezogen	-
Akteur:innen	Bibliotheksleitung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige des Wachstums des Bibliotheksbestandes nach einzelnen <i>RVK</i> -Systematikstellen.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen lassen sich auf der <i>GUI</i> das Bestandswachstum nach <i>RVK</i> -Systematikstellen anzeigen.	<p>Das System filtert die betreffenden Datensätze für die <i>RVK</i>-Systematikstellen. Das System zeigt diese Datensätze mit Datenvisualisierungen an.</p> <p>Das System zeigt die Anzahl der Titel pro <i>RVK</i>-Systematikstelle an.</p> <p>Das System zeigt darüberhinaus die Gesamtzahl der Titel nach Jahren an.</p> <p>Das System zeigt darüberhinaus die Gesamtzahl der Titel und die Zahl der <i>RVK</i>-Systematikstellen an.</p> <p>Das System zeigt darüberhinaus die Anzahl der Titel nach Medienart für jede <i>RVK</i>-Systematikstelle pro Jahr an.</p> <p>Das System zeigt darüberhinaus die Entwicklung in der Verteilung der Titel in den <i>RVK</i>-Systematikstellen über den Gesamtzeitraum an.</p>
Anforderungen	R1, R4, F2-F4, F10-F17, NF9, NF14-NF18	-

Tabelle 4.10: Anwendungsfall 5 - Bestandswachstum

Anwendungsfall 6

	Beschreibung	Systemverhalten
Titel	Budget- und Umsatzübersicht	-
Evaluationstyp	Sammlungsbezogen	-
Akteur:innen	Bibliotheksführung, Bibliotheksmitarbeiter:innen	-
Ziel	Anzeige der Budget- und Umsatzübersicht für den Gesamtzeitraum und das laufende Jahr.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	Das System ist im Betrieb.
Inhalt	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen können sich den Lieferanten auswählen und den Gesamtumsatz und den Umsatz pro Jahr ansehen.	Das System filtert die betreffenden Datensätze für den Lieferanten. Das System zeigt diese Datensätze mit Datenvisualisierungen an.
	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen können sich die Budgetübersicht nach Kostenstelle auswählen und Gesamtbudget per Kostenstelle und pro Jahr ansehen.	Das System filtert die betreffenden Datensätze für die Kostenstellen. Das System zeigt diese Datensätze mit Datenvisualisierungen an.
	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen können sich den Verlauf des Budgets und der Umsätze über den Gesamtzeitraum und über das laufende Jahr anzeigen lassen.	Das System zeigt das Budget und den Umsatz für den Gesamtzeitraum und für das laufende Jahr an.
		Das System zeigt darüberhinaus Entwicklung für das Budget und den Umsatz über den Gesamtzeitraum an.
Anforderungen	R1, R4, F2, F3, F10-F17, NF9, NF14-NF18	-

Tabelle 4.11: Anwendungsfall 6 - Budget- und Umsatzübersicht

4 Konzeption einer Lösung

Anwendungsfall 7

	Beschreibung	Systemverhalten
Titel	Standardbericht	-
Evaluationstyp	-	-
Akteur:innen	Bibliotheksleitung, Bibliotheksmitarbeiter:innen, Institutsleitung	-
Ziel	Generierung eines Standardberichts mit den relevanten <i>KPI</i> der Bibliothek zur Vorlage bei der Institutsleitung.	Das System generiert eine Anzeige mit den jeweiligen Parametern.
Vorbedingungen	Die Bibliotheksleitung und die Bibliotheksmitarbeiter:innen haben Zugriff auf das System.	
Inhalt	Die Bibliotheksmitarbeiter:innen lösen ein Skript zur Generierung des Standardberichtes aus.	Das System greift auf Bilder und Texte an einem definierten Speicherort zu.
		Das System generiert aus diesen Bildern und Texten automatisch ein PDF-Dokument.
		Das System öffnet das PDF-Dokument automatisch.
	Die Bibliotheksleitung oder die Bibliotheksmitarbeiter:innen speichern das PDF-Dokument an einem Speicherort.	Speicherung des PDF-Dokument systemunabhängig.
Anforderungen	Die Bibliotheksleitung oder die Bibliotheksmitarbeiter:innen verteilen das PDF-Dokument an die Institutsleitung.	-
	R2, F18-F21, NF14-NF18	-

Tabelle 4.12: Anwendungsfall 7 - Standardbericht

5 DISKUSSION DER UMSETZUNG

Im folgendem Kapitel wird das Design des Prototypen besprochen und die Teilsysteme vorgestellt und skizziert wie die Teilsysteme funktionieren. Dabei wird exemplarisch anhand von einzelnen Methoden der Klassen oder Funktionen beschrieben wie das funktioniert. Im Quellcode selber wird ausführlich jede Methode beschrieben.

5.1 IMPLEMENTIERUNG

5.1.1 TECHNISCHE DETAILS DER IMPLEMENTIERUNG

Das Proof-of-Concepts wurde mittels der Programmiersprache Python umgesetzt. Python ist eine höhere Programmiersprache. Sie ist weitverbreitet [vgl. [Lou21](#)] und besitzt eine einfache Syntax. Besonderheiten von Python sind der Verzicht auf geschweifte Klammern und Interpunktionen nach Anweisungen. Die Anweisungen sind dagegen durch Einrückungen strukturiert und nicht durch öffnende und schließende Klammern getrennt. Des Weiteren zeichnet sich Python durch eine dynamische Typisierung aus und ist dadurch für Skripte als auch für die schnelle Entwicklung von Anwendungen geeignet. Python erlaubt die Aufteilung von Programmen in Modulen, die in anderen Python-Programmen wiederverwendet werden können [vgl. [Pyt21a](#)]. Sie ist für alle wichtigen Plattformen frei verfügbar und besitzt eine große Dokumentation [vgl. [Pyt21c](#)]. Wie andere Programmiersprachen besitzt Python eine umfangreiche Standardbibliothek. Dazu gibt es noch eine Vielzahl von Modulen, Programmen und Werkzeugen von Drittanbietern [vgl. [Pyt21b](#)].

5 Diskussion der Umsetzung

Genutzt wurde für das Projekt insbesondere die Python-Bibliothek *pandas*, die eine Open-Source-Bibliothek für Datenanalyse und Datenmanipulation darstellt [vgl. Pan21]. Wichtige Konzepte dieser Bibliothek sind der *pandas* Dataframes und die *pandas* Series, die *pandas* Objekte sind. Der *pandas* Dataframe ist eine zweidimensionale tabellarische Datenstruktur mit beschrifteten Achsen (Spalten und Reihen). In ihn werden die Daten unter anderem aus csv- oder xlsx-Dateien über *pandas* Funktionen geladen. Eine *pandas* Series kann Bestandteil eines Dataframes sein. Der Datentyp entspricht eines eindimensionalen Arrays. [Abbildung 5.1](#) zeigt anhand der Umsatzdaten die ersten fünf Reihen eines *pandas* Dataframe und einer *pandas* Series desselben Dataframes. Zugriff auf die Series erfolgt über den Spaltenkopf 'Lieferant Abk.' der als ein Schlüssel zur Series dient. Sowohl der Dataframe als auch die Series haben einen Index, über den auf die einzelnen Werte zugegriffen werden kann.

	Lieferant	Umsatz (EUR)	Datum	Lieferant Abk.		
0	Antiquariat allgemein	2679.35	2018-12-01	Antiquariat	0	Antiquariat
1	Autorenbuchhandlung Marx & Co GmbH	13020.71	2018-12-01	Autorenbuchhandlung Marx	1	Autorenbuchhandlung Marx
2	Buchhandlung Erhardt & Kotitschke GbR	1177.78	2018-12-01	Buchhandlung Erhardt	2	Buchhandlung Erhardt
3	COMPUTEC MEDIA GmbH	77.20	2018-12-01	COMPUTEC MEDIA	3	COMPUTEC MEDIA
4	Casalini Libri	120.00	2018-12-01	Casalini Libri	4	Casalini Libri

Abbildung 5.1: *pandas* Dataframe und *pandas* Series

Auf dem Dataframe, der als sogenannter Container für die Series dient, können verschiedene Operationen der Datenanalyse und -manipulation erfolgen. Ähnlich wie bei Abfragen einer Datenbank lassen sich verschiedene Funktionen wie `sort()`- oder `groupby()`-Funktionen in Zusammenhang mit Aggregatfunktionen wie `mean()`, `median()` oder `sum()` auf den Series durchführen. Auch gibt es eine Vielzahl von Funktionen, um die Daten rasch zu explorieren. Die Werte der Zellen können verschiedene Datentypen annehmen. Zum Beispiel können die Datumswerte der Spalte 'Datum' in `datetimeformat`-Objekte umgewandelt werden.

In Verbindung mit pandas ist Python neben R und Matlab im wissenschaftlichen Kontext für Data-Science-Projekte sehr beliebt.

Es gibt eine Vielzahl an Bibliotheken für Python, die für die Entwicklung von interaktiven Datenvisualisierungen verfügbar sind. Zu nennen wären Bokeh [vgl. [Van21](#)], Altair [vgl. [Alt21](#)] und die Graphic Libraries von Plotly [vgl. [Plo21e](#)]. Für das Projekt wurde sich für Plotly Express entschieden und dieses hauptsächlich genutzt. Plotly Express ist eine Graphische Bibliothek für Python, R und Javascript. Es ist eine Weiterentwicklung der Bibliothek Plotly Graphic Objects. Plotly Express besitzt eine einfachere Syntax bei fast annähernder Feature-Gleichheit mit Plotly Graphic Objects. So können mit wenigen Zeilen Python-Code interaktive Grafiken erzeugt werden. An interaktiven Basisfunktionalitäten bietet Plotly Express zum Beispiel Hover-Informationen der Datenpunkte, Zoom-In und Zoom-Out-Möglichkeiten, das Aus- und Abwählen von Balken oder Linien in den entsprechenden Diagrammen. Plotly Express kann pandas Dataframes oder pandas Series als Datenobjekte erwarten und verarbeiten. Dabei können die Spaltenköpfe die Achsen darstellen und die einzelnen Werte der Spalten die Datenpunkte bilden. Die Rendering-Engine für die Diagramme beruht auf dem JavaScript Framework D3.js. Plotly Express ist frei verfügbar. In Kombination mit der Bibliothek Dash für die Entwicklung von interaktiven Dashboards lässt sich Plotly Express gut anwenden, da beide von der selben Firma entwickelt werden. Mit der Bibliothek Dash wurde das Dashboard realisiert.

Dash baut auf den Technologien Flask, Plotly.js und React.js auf. Das Framework ermöglicht die Erstellung interaktiver Webapplikationen oder 'Analytical Applications' in Python ohne Javascript programmieren zu müssen [vgl. [Plo21c](#)]. Es gibt verschiedene Dash-Komponenten, die unterschiedliche Funktionen erfüllen. `Dash_html_components` stellen Klassen für alle HTML-Tags bereit. Die Schlüsselwortargumente dieser Komponente beschreiben HTML-Attribute wie `style`, `className` und `id` [vgl. [Plo21d](#)]. Andere Komponenten sind `dash_core_components` oder `dash_dependencies` [vgl. [Plo21b](#)]. Während die `dash_core_components` Steuerelemente und Graphen erzeugen, regeln die `dash_dependencies`

über Callback-Decorator-Funktionen die Interaktion zwischen den einzelnen Komponenten. So kann zum Beispiel in der Dash-WebApplikation das Verhalten eines Diagramms von den Werten eines Dropdown-Menü gesteuert werden. Die zugehörige Callback-Decorator-Funktion wird in diesem Falle automatisch von Dash aufgerufen, sobald sich der Wert des Dropdown-Menüs ändert. Dabei bewältigt Dash alle Javascript-Anforderungen im Front- und im Backend [vgl. Plo21a]. Dash_bootstrap_components ist eine weitere Bibliothek, die zum Einsatz in diesem Projekt kommt. Sie unterstützt die grafische Umsetzung des Dashboardes [vgl. Fac21].

Tabelle 5.1 zeigt nochmals einen kurzen Überblick über die Versionsnummern der genutzten Programmiersprache und der hauptsächlich genutzten Bibliotheken sowie deren Open-Source Lizenzen.

Name	Version	Lizenz	Webseite
Python	3.7.9	Open Source (PSF)	https://docs.python.org/3.7/
pandas	1.1.5	3-Clause-BSD-License	https://pandas.pydata.org/pandas-docs/version/1.1.5/
Plotly	4.14.14	MIT-License	https://plotly.com/python/
Dash	1.18.1	MIT-License	https://dash.plotly.com/

Tabelle 5.1: Liste der grundlegenden Programmiersprache und Frameworks

5.1.2 SYSTEMARCHITEKTUR

Das System teilt sich in drei Teilsysteme auf, die im Unterabschnitt 5.1.3 näher beschrieben werden. Tabelle 5.2 zeigt die drei Teilsysteme mit einer Kurzbeschreibung der Hauptaufgabe. Teilsystem 1 ist für die erste Bereinigung der Daten und den Import der Daten in ein einheitliches Datenformat verantwortlich. Teilsystem 2 greift auf den Speicherort der importierten Dateien zu und bereitet die Daten für die Anzeige im Dashboard vor, indem es mit pandas Daten vorfiltert, gruppiert und Berechnungen auf den Daten ausführt. Teilsystem 3 ist für das Layout des Dashboards, die Erstellung der Diagramme, die

Anordnung der Diagramme, die Bereitstellung von Interaktionen mit dem Dashboard im Frontend verantwortlich. Teilsystem 1 und Teilsystem 2 wurden versucht objektorientiert zu programmieren, da hier der Anspruch bestand, den Programmcode für verschiedene Daten wiederzuverwenden. Teilsystem 3 besteht vorerst nur aus Funktionen, die die Anzeige im Frontend ermöglichen.

Teilsystem Name		Hauptaufgabe
1	Import	Import und erste Bereinigung der Daten aus heterogenen Datenquellen.
2	Datenbearbeitung	Aufbereitung der Daten für die graphische Darstellung im Dashboard.
3	Darstellung	Umwandlung der Daten in Datenvisualisierungen, etc. und Anzeige im Dashboard.

Tabelle 5.2: Liste der Teilsysteme

5.1.3 TEILSYSTEME

Teilsystem 1 Import

Das Teilsystem Import ist verantwortlich für den Import der Daten im Rohformat aus vordefinierten Importverzeichnissen in vordefinierte Zielverzeichnisse. Das Ziel ist einerseits die Daten automatisch ohne Informationsverlust zu importieren und andererseits diese mit notwendigen Daten für die weitere Analyse anzureichern und erste Bereinigungen der Daten durchzuführen. Die Daten werden zum Abschluss in csv-Format abgespeichert. Mit dem Teilsystem soll eine einheitliche Datengrundlage für die spätere Bearbeitung und Darstellung garantiert werden. Für das csv-Format wurde sich aufgrund seiner flachen und einfachen Struktur, der guten Lesbarkeit und seiner weiten Verbreitung entschieden. Das Teilsystem Import besteht aus vier Klassen. Die [Abbildung 5.2](#) zeigt die einzelnen Klassen mit ihren Methoden. Die Klassen im Teilsystem Import sind auf die Daten aus fremden Quellen zugeschnitten (Budget Umsatz, Neuerwerbungslisten, Ausleihe (Anwendungsfall 2, 4-6)). Dennoch können mit ihnen auch die bibliotheksinternen Daten wie Lesesaalnutzung (Anwendungsfall 3) bearbeitet werden.

5 Diskussion der Umsetzung

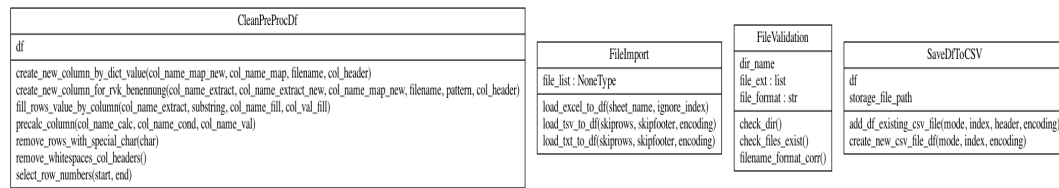


Abbildung 5.2: Klassendiagramm - Teilsystem Import

Instantiiert werden die einzelnen Klassen für die jeweiligen konkreten bibliothekarischen Daten in eigenen Skripten. So gibt es Skripte für Budget, Umsatz, Ausleihe und Bestand/Neuerwerbungen. Diese verwenden für die Daten passende Methoden.

Den Ablauf des Teilsystems zeigt schematisch die [Abbildung 5.3](#).

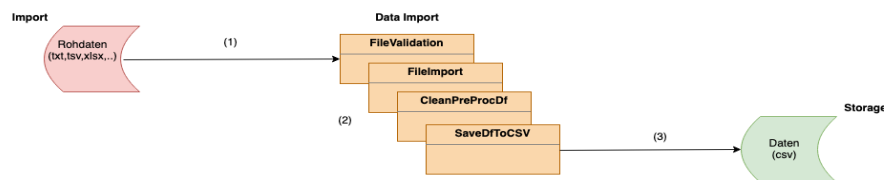


Abbildung 5.3: Datenfluss - Teilsystem Import

(1) Für den ersten Schritt sind die Klassen `FileValidation` und `FileImport` verantwortlich. Die Dateien werden aus einem vorher definierten lokalen Verzeichnis in ein pandas Dataframe geladen. Dabei wird mit Methoden der Klasse `FileValidation` sichergestellt, dass sowohl das Verzeichnis als auch die Dateien existieren. Des Weiteren wird sichergestellt, dass die Dateinamen einem definierten semantischen Format wie dem Datumsformat `YYYY_MM_DD` und einem Dateiformat wie `txt`, `xlsx` oder `tsv` entsprechen. Wenn die Daten nicht diesen Vorgaben entsprechen, werden sie nicht importiert. Da die Daten unterschiedlich aufgebaut und in unterschiedlichen Dateiformaten vorliegen werden beim Import in den pandas Dataframe jeweils verschiedene Methoden angewandt.

Beim Ladeprozess der Dateien in den pandas Dataframe wird mit den Methoden `load_txt_to_df` oder `load_tsv_to_df` der Klasse `FileImport` der Dateiname von der Datei extrahiert und in einer neu geschaffenen Spalte des Dataframes im Datumsformat `YYYY-`

MM-DD gespeichert. Anhand den Werten dieser Spalte werden später verschiedene Datenanalysen und Datenvisualisierungen vollzogen.⁸ Verantwortlich für die Extrahierung des Dateinamens ist die in `utils.py` ausgelagerte Funktion `date_from_filename`, die den Dateinamen als Argument entgegennimmt.⁹

(2) Das geladene pandas Dataframe wird im zweiten Schritt durch die Klasse `CleanPreProcDf` aufgenommen und durch verschiedene Methoden dieser Klasse manipuliert. Beispielhaft ist hier die Methode `create_new_column_for_rvk_benennung` zu nennen, die aus der Spalte der Signatur der Neuerwerbungen die Hauptklassen der *Regensburger Verbundklassifikation* extrahiert und in einer neuen Spalte speichert. Dieser Methode liegt eine csv-Datei der *RVK* zu Grunde, die mit einem Skript aus der *RVK*-XML-Datei entstanden ist. Manuell ergänzt wurde diese noch um selbstgeschaffene Hauptklassen der Institutsbibliothek. Die Methode wird sowohl bei den Neuerwerbungs- und Ausleihdaten angewendet. Ebenso entstehen neue Spalten bei Umsatz- und Budgetdaten. In Bezug auf die Darstellung der Daten im Dashboard werden hier die Namen der Lieferanten und Kostenstellen in lesbarer Form in einer neuen Spalte gespeichert. Des Weiteren gibt es in der Klasse Methoden, die für die Entfernung von Reihen mit bestimmten syntaktischen Zeichen wie Bindestrichen verantwortlich sind oder die unnötige Leerzeichen in Spaltenköpfen löschen und mit einem Leerzeichen ersetzen. Ferner werden erste Berechnungen auf den Daten ausgeführt, um valide Daten für die Auswertung zu haben. Dies betrifft die Ausleihdaten und wird durch die Methode `precalc_column` realisiert.¹⁰

(3) Nachdem Transformationsprozess wird das veränderte pandas Dataframe als csv-Datei in einem vorher definierten Zielordner gespeichert. Verantwortlich ist dabei die Klasse `SaveDfToCSV` mit den zwei Methoden `add_df_existing_csv_file` `create_new_csv_file_df`.

⁸ Dieses Verfahren wird bei den Daten ausgeführt, die einer zusätzlichen Datumsspalte bedürfen.

⁹ In der Datei `utils.py` sind noch andere Funktionen als stand-alone-functions für den Datenimport und Datenbearbeitung gruppiert, da diese das Objekt `self` nicht verändern.

¹⁰ Im Arbeitsprozess der Medienschließung validiert die Bibliothek die RFID-Etiketten der Medien durch einmalige Ausleihe der Medien am Selbstverbucher. Deswegen wird die Anzahl der Ausleihe pro Medium um eins reduziert.

5 Diskussion der Umsetzung

Die je nach dem mit dem Dataframe eine neue csv-Datei erstellen oder an eine bereits vorhandene csv-Datei den Dataframe anhängen. [Abbildung 5.4](#) zeigt anhand einer Umsatz-Datei, die im txt-Format gespeichert ist, die Transformation von diesem Format in ein csv-Format. Für die lesbare Darstellung wird die csv-Datei in einem Tabellenformat angezeigt.

Umsatzübersicht Bibliothek MPI fuer Empirische Aesthetik			
Stand: Nov 1 2020 6:16AM			

Lieferant	Umsatz (EUR)		

Antiquariat allgemein	0.00		
Autorenbuchhandlung Marx & Co GmbH		0.00	
COMPUTEC MEDIA GmbH		0.00	
Deutscher Hochschulverband	0.00		
EBSCO Information Services GmbH		0.00	
EBSCO Information Services GmbH Buch			0.00
Erasmus Boekhandel bv		0.00	
Ersatzlieferant	0.00		
Felix Meiner Verlag GmbH		0.00	
Heinemann Verlag Shop		0.00	
Karl-Marx-Buchhandlung GmbH		0.00	
Massmann Internationale Buchhandlung GmbH			0.00
Missing Link Versandbuchhandlung eG		0.00	
Presse Verlagsgesellschaft fr Zeitschriften und neue Medien mbH			0.00
Schweitzer Fachinformationen		0.00	
Springer	0.00		
Texte zur Kunst GmbH & Co. KG			0.00
Verlagsgruppe Hthig Jehle Rehm			0.00
ebooks	0.00		
	Umsatz (EUR)		

Summe	0.00		

1	Lieferant	Umsatz (EUR)	Datum	Lieferant Abk.
2	Antiquariat allgemein	0.00	2020-10-01	Antiquariat
3	Autorenbuchhandlung Marx & Co GmbH	0.00	2020-10-01	Autorenbuchhandlung Marx
4	COMPUTEC MEDIA GmbH	0.00	2020-10-01	COMPUTEC MEDIA
5	Deutscher Hochschulverband	0.00	2020-10-01	Deutscher Hochschulverband
6	EBSCO Information Services GmbH	0.00	2020-10-01	EBSCO Zeitschriften
7	EBSCO Information Services GmbH Buch	0.00	2020-10-01	EBSCO Buch
8	Erasmus Boekhandel bv	0.00	2020-10-01	Erasmus Boekhandel
9	Ersatzlieferant	0.00	2020-10-01	Ersatzlieferant
10	Felix Meiner Verlag GmbH	0.00	2020-10-01	Felix Meiner Verlag
11	Heinemann Verlag Shop	0.00	2020-10-01	Heinemann Verlag
12	Karl-Marx-Buchhandlung GmbH	0.00	2020-10-01	Karl-Marx-Buchhandlung
13	Massmann Internationale Buchhandlung GmbH	0.00	2020-10-01	Massmann
14	Missing Link Versandbuchhandlung eG	0.00	2020-10-01	Missing Link
15	Presse Verlagsgesellschaft fr Zeitschriften und neue Medien mbH	0.00	2020-10-01	Presse Verlagsgesellschaft
16	Schweitzer Fachinformationen	0.00	2020-10-01	Schweitzer
17	Springer	0.00	2020-10-01	Springer
18	Texte zur Kunst GmbH & Co. KG	0.00	2020-10-01	Texte zur Kunst
19	Verlagsgruppe Hthig Jehle Rehm	0.00	2020-10-01	Hüthig Jehle Rehm
20	ebooks	0.00	2020-10-01	ebooks

Abbildung 5.4: Monatliche Umsatzübersicht nach Ablauf Teilsystem Import

Teilsystem 2 Datenbearbeitung

Das Teilsystem 2 hat als Ziel die Daten für die Darstellung im Dashboard vorzubereiten. Der weitere Zweck dieses Teilsystem ist es, die Vorbereitung der Daten von der Erstellung der Datenvisualisierungen zu trennen. Ein Grund hierfür ist, den Programmcode im Teilsystem 3 nicht mit dem Programmcode der Datenmanipulation zu überfrachten. Ebenfalls haben Überlegungen der Latenzzeit eine Rolle gespielt. Deswegen werden in dem Teilsystem Datenbearbeitung Daten mit pandas so bearbeitet, dass entweder Teilmengen der eigentlichen Daten oder Ergebnisse mathematischer Operationen durch das Teilsystem 3 zu Datenvisualisierungen nur noch weiterverarbeitet werden müssen. Die Vorbereitung der Daten umfasst einzelne Berechnungen wie `mean()` oder `sum()`. Ferner werden im Teil-

system 2 die Sortierung, Gruppierung, Filterung der Daten nach bestimmten Aspekten vollzogen, die in den Anwendungsfällen formuliert sind.

Für das Teilsystem Datenbearbeitung ist das Modul `data_prep` zentral. Die Klassenstruktur dieses Moduls ist eine Basisklasse von der vier Kindklassen erben. Es gibt Kindklassen für Umsatz und Budget-Daten, für Bestands- und Neuerwerbungsdaten, für Daten der Ausleihe und der Lesesaalnutzung. Aufgrund der ähnlichen Datenstruktur der Umsatz- und Budgetdaten genügt eine Klasse für beide. Der Gesamtbestand und die monatlichen Neuerwerbungen berufen sich auf denselben Datenbestand, so dass hier ebenfalls auf eine Aufteilung auf mehrere Klassen verzichtet wurde. Für jede Kindklasse wurden Methoden geschrieben, die auf den spezifischen Bibliotheksdaten verschiedene Manipulationen ausführen. Die Ergebnisse werden von jeder Methode zurückgegeben. [Abbildung 5.5](#) zeigt die Basisklasse und die vier Kindklassen mit ihren Methoden.

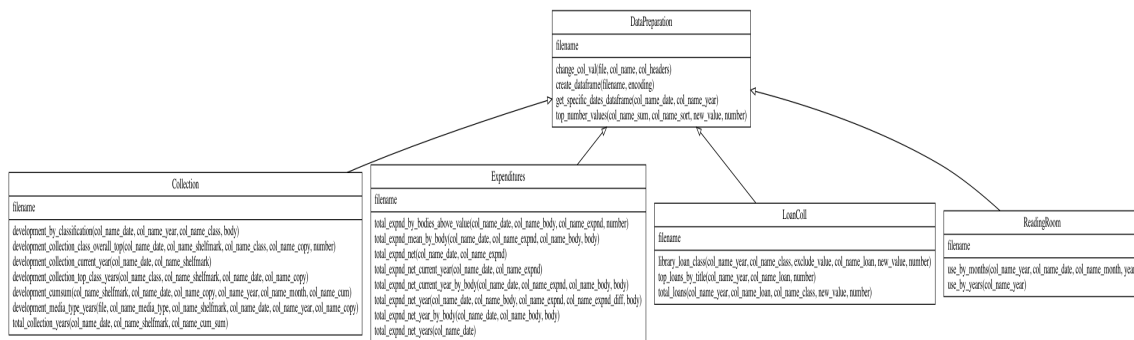


Abbildung 5.5: Klassendiagramm - Teilsystem Datenbearbeitung

Der Ablauf innerhalb des Teilsystems besteht aus zwei Schritten.

(1) Das Laden der einzelnen csv-Dateien aus dem Zielverzeichnis wird durch die Methode `create_dataframe` der Basisklasse geregelt. Es wird noch einmal sichergestellt, dass die Daten in einem Zeichenformat wie utf-8 in das pandas Dataframe geladen werden. Beim Aufrufen des Konstruktor beziehungsweise der `init`-Methode der Klasse wird die Methode `create_dataframe` automatisch mit aufgerufen. Da die Methode und die Konstruktor-Properties der Basisklasse mit vererbt werden, kann in den Kindklassen das

Objekt ebenfalls mit dem pandas Dataframe initiiert werden. Die Objekte werden im nächsten Schritt durch die spezifischen Methoden der Kindklassen bearbeitet.

(2) Das Ergebnis der Manipulation der pandas Dataframes ist auf die Darstellung der Daten im Dashboard ausgerichtet. Als Input der einzelnen Methoden der Kindklassen werden verschiedene Parameter verlangt, mit denen dann der Dataframe entweder manipuliert oder auf ihm Berechnungen ausgeführt werden können. Um die Datenpreparation zu erreichen, können verschiedene Transformationsschritte auf den Daten ausgeführt werden. Die Rückgabewerte der Methoden sind veränderte pandas Dataframes, pandas Series oder ein Ergebnis einer mathematischen Operation.

Anhand der folgenden Methode soll der letzte Schritt verdeutlicht werden. Die Methode `total_expnd_net_current_year` der Expenditures-Klasse bestimmt im Allgemeinen die Summe von Werten einer Spalte, gefiltert nach einem Wert einer anderen Spalte eines pandas Dataframes. Mit dieser Methode soll so zum Beispiel konkret der Gesamtumsatz des laufenden Jahres berechnet werden.

```
def total_expnd_net_current_year(self, col_name_date, col_name_expnd):  
    ...  
    self._date_max = self._df[col_name_date].max()  
    self._df = self._df.set_index(col_name_date)  
    self._df = self._df.loc[self._date_max]  
  
    return self._df[col_name_expnd].sum()
```

Quellcode 5.1: Beispiel Methode Exenditures class

Als Input erwartet die Methode in der Methodensignatur zwei Spaltennamen als Parameter: den Namen der Datumsspalte `col_name_date` nach der gefiltert wird und den Namen der Umsatz-Spalte `col_name_expnd` auf der die Berechnung stattfindet. Der Transformationsprozess im Methodenkörper teilt sich in drei Schritte auf. Da die monatlichen Umsatzdaten im Jahr als akkumulierte Daten vorliegen, interessieren nur die letzten vor-

handenen Datensätze des laufenden Jahres. Deswegen wird nach diesen gefiltert und ein Dataframe von diesen erstellt. Dementsprechend wird mit der pandas-Funktion `max()` der maximale Wert in der Datumsspalte zunächst bestimmt und der Variable `_date_max` zugewiesen. Danach wird die Datumsspalte als Index gesetzt. Im dritten Schritt wird der Dataframe mit den Reihen, die der Variable `_date_max` entsprechen, erstellt. Dies geschieht mit der pandas-Funktion `.loc`, die auf den Index der Reihen des Dataframes zugreift. Zum Schluss wird auf Basis der Umsatz-Spalte des Dataframes die Summe mit der pandas-Funktion `sum()` berechnet und als Rückgabewert zurückgeliefert. Der Rückgabewert kann nun vom Teilsystem 3 Darstellung weiterverarbeitet werden.

Teilsystem 3 Darstellung

Für die Erschaffung des Dashboards mit seinen Datenvisualisierungen und Interaktionen ist das Teilsystem Darstellung verantwortlich. In ihm werden die Ergebnisse des Teilsystems 2 zu Datenvisualisierungen verarbeitet und die Dashboard-Logik bereitgestellt. Dazu werden die Bibliothek Plotly Express für die Datenvisualisierungen und die Bibliothek Dash zur Umsetzung des Dashboards genutzt.

Die Struktur der Dashboard-App entspricht einem Multi-App-Dashboard. Aufgrund der Vielzahl an Datenvisualisierungen wurde sich gegen eine Single-Page-Lösung entschieden. Deswegen besteht das Dashboard aus drei einzelnen Tabs.

Es gibt verschiedene Möglichkeiten Multi-App-Dashboard-Projekte zu strukturieren. Für das vorliegende Projekt wurde sich für die in [Abbildung 5.6](#) gezeigte Struktur entschieden.

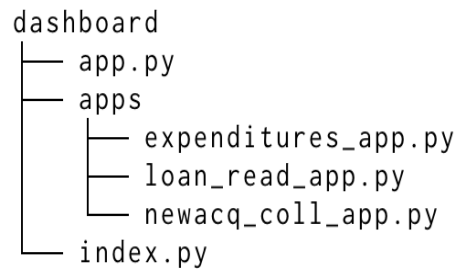


Abbildung 5.6: Struktur Dash Multi-Page App

Dies entspricht einer der Möglichkeiten, wie sie auf der Webseite von Dash für Multi-App-Projekte vorgeschlagen werden [vgl. [Plo21f](#)].

Die einzelnen Tabs wurden versucht inhaltlich um bibliothekarische Tätigkeiten zu gruppieren. So werden in der `expenditures_app.py` Datenvisualisierungen erstellt, die Budget- und den Umsatzdaten visualisieren, während mit Hilfe der `loan_read_app.py` Ausleih- und Lesesaalnutzungsdaten dargestellt werden. Mit der `newacq_coll_app.py` wird ermöglicht Daten aus dem Bereich der Bestandsentwicklung und der Ausleihe Daten zu präsentieren.

Für jeden Tab gibt es eine separate Datei. Jede einzelne Tab-Datei besteht einerseits aus mehreren Funktionen für die Erschaffung von Datenvisualisierungen der einzelnen bibliothekarischen Daten. Innerhalb dieser Funktionen werden mit den Funktionen der Bibliothek Plotly Express `Plotly Graph Object Figure` geschaffen. Andererseits bestehen die Dateien aus mehreren verschiedenen Funktionen für die Dash-Komponenten wie zum Beispiel Dropdown-Menüs, Cards oder Diagrammen. Die Funktionen der Dash-Komponenten binden die Funktionen der Datenvisualisierung so ein, dass die `Plotly Graph Object Figure` im Dashboard zur Anzeige gebracht werden können. Weiterhin sind die Datenvisualisierungs-Funktionen zum Teil mit Decorator-Callback-Funktionen verknüpft, die es ermöglichen mit dem Dashboard zu interagieren. So wurde in zwei Tab-Dateien die Logik von einfachen Callbacks implementiert. Weiterhin enthalten die Dateien jeweils eine Layoutfunktion, die das gesamte Layout des Tabs bündelt.

Neben den Funktionen für die Dash-Komponenten und der Datenvisualisierungen, werden in den tab-Dateien noch die Objekte aus dem Modul `data_prep` instantiiert und die Methoden der Kindklassen aus diesem Modul auf diese angewendet. In der Regel geschieht dies ausserhalb der Datenvisualisierung-Funktionen. Für die Callback-Funktionalität werden aber die Objekte und die Methoden des Moduls `data_prep` innerhalb der Datenvisualisierung-Funktionen aufgerufen.

Schematisch zeigt die [Abbildung 5.7](#) den Ablauf in den Tab-Dateien ohne die callback-Funktionen. Die Objekte werden in Plotly Graph Objects umgewandelt. Diese wiederum werden in Dash-Objekte umgewandelt und werden in einem Layout für das Tab zusammengefasst.

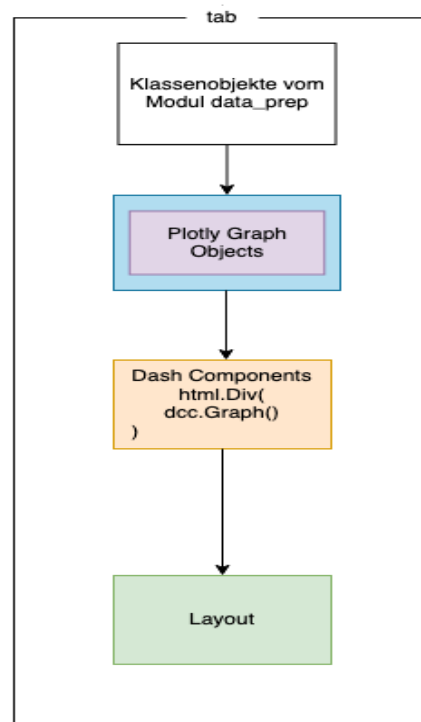


Abbildung 5.7: Ablauf Tab

Dabei sieht die grundsätzliche Logik der Funktionsaufrufe in den Tab-Dateien ohne die Callback-Funktionen wie folgt aus: [Abbildung 5.8](#) aus:

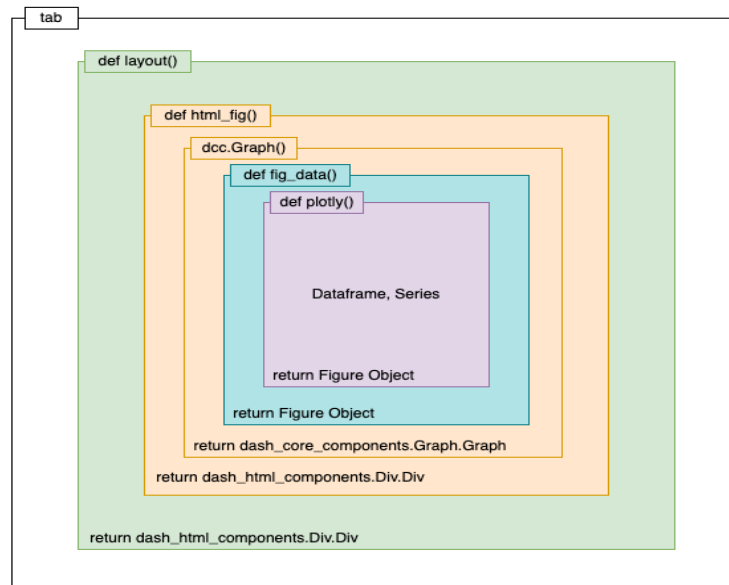


Abbildung 5.8: Funktionsaufrufe Tab

In der Plotly-Funktion, die das `Plotly Graph Object Figure` erzeugt, werden die durch die Kindklassen bearbeiteten Objekte (`Dataframe`, `Series`) als Parameter entgegengenommen. Zusätzlich müssen noch weitere Parameter übergeben werden, die für die jeweiligen Diagrammtypen wichtig sind. Andere Parameter, die das Layout festlegen, sind dagegen optional.

Im folgenden Beispiel wird in der Funktion `fig_total_expnd` aus `expenditures_app.py` ein horizontal gekipptes Balkendiagramm-Objekt mit dem pandas `Dataframe` `df_total_expnd`, der x-Achse 'Umsatz (EUR)' und y-Achse 'Datum' durch den Aufruf der Plotly-Funktion `bar` erzeugt und der Variable `fig` zugewiesen.

```
def fig_total_expnd():
    ...
    fig = px.bar(df_total_expnd,
                 x='Umsatz (EUR)',
                 y='Datum',
                 orientation='h',
                 ...
```

```
)
...
```

Quellcode 5.2: Funktion `fig_total_expnd()` Auszug 1

Mit weiteren Funktionen von Plotly Express können `Plotly Graph Object Figure` bearbeitet werden. So kann durch die `update_layout`-Funktion unter anderem der Achsenvitel, Höhe und Breite der Datenvisualisierung festgelegt werden. Als Rückgabewert der Funktionen `fig_data` werden die `Plotly Graph Object Figures` zurückgegeben.

```
def fig_total_expnd():
    ...
    fig.update_layout(title_x=0.5,
                      xaxis_title='Umsatz in Euro',
                      yaxis_title='Jahr',
                      height=500)
    fig.update_xaxes(nticks=20)
    return fig
```

Quellcode 5.3: `fig_total_expnd()` Auszug 2

In dem Beispiel werden der Titel der x- und y-Achse, sowie die Größe des Diagramm-Objektes festgelegt. Schließlich wird noch die x-Achse mit der Funktion `update_xaxes` skaliert.

Die `fig_data`-Funktionen werden durch die Graph-Komponente der `dash_core_components` innerhalb der `html_fig`-Funktionen aufgerufen. Diese Komponente ist für die Umsetzung der interaktiven Datenvisualisierungen zuständig. Ebenfalls kann von der Graph-Komponente eine `id` als Parameter entgegengenommen werden. Diese ist unter anderem wichtig für die Callback-Funktionen.

```
def html_fig_total_expnd():
    ...
```

```
return html.Div(  
    [  
        html.Div(  
            [  
                dcc.Graph(  
                    id='gesamtumsatz',  
                    figure=fig_total_expnd()  
                ),  
            ], className="six columns chart_div", style={'margin-top': '20px', 'margin  
-left': '10px'}  
        ),  
    ]  
)
```

Quellcode 5.4: `html_fig_total_expnd()`

Die `html_fig`-Funktionen werden dann in einer Layoutfunktion eingebunden, die das Layout für den Tab bündelt und als Rückgabewert ebenso wie die `html_fig`-Funktionen eine Div-Komponente der `dash_html_components` returniert.

Die Callback-Funktionen wurden für zwei Dropdown-Menüs in zwei Tab-Dateien implementiert. Der schematische Ablauf in der Tab-Datei mit den Callback-Funktionen zeigt die [Abbildung 5.10](#). Der Callback ist als Dekorator für eine Funktion implementiert. In der Implementierung wird ein Callback ausgelöst, wenn ein Wert über das Dropdown ausgewählt wird. Der Callback übergibt diesen Wert an die Funktion `update_output_div()`. Die Funktion returniert unter anderem `fig_data()`-Funktionen mit diesem Wert als Argument. Diese Funktionen übergeben den Methoden der Kindklassen aus dem Modul `dataprep` das Argument. Diese erzeugen einen Dataframe basierend aus den übergebenden Argument von dem Objekt und geben mit Hilfe der `plotly`-Funktionen eine `Plotly Graph Object Figure` der gefilterten Daten zurück. Der Callback übergibt dann die zurückgegebene Figur an die im Output angegebene Komponente.

In der `expenditures_app.py` sind zum Beispiel jeweils zwei Diagramme und Zahlenwerte von den Werten einer Dropdown-Liste abhängig.

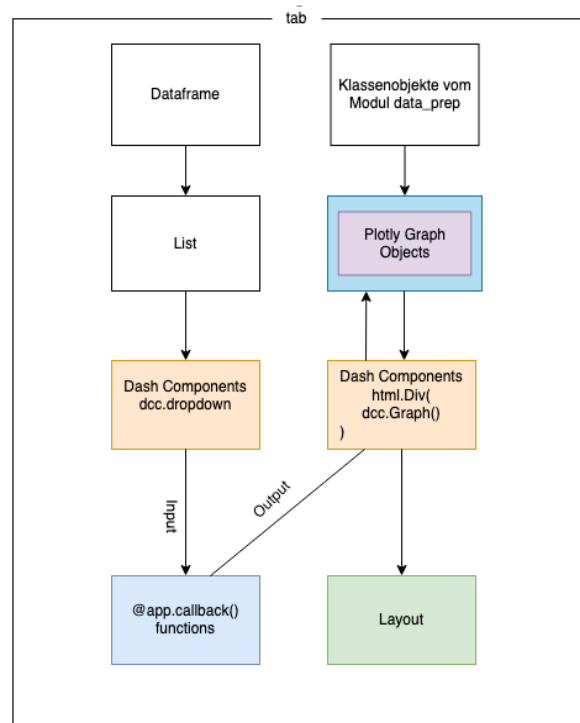


Abbildung 5.9: Ablauf Tab mit Callback

Das Dropdown-Menü wird aus unique Werten einer Dataframespalte erzeugt. Diese werden dann mit der Dropdown-component von der dcc zu einem Dropdown-Element umgewandelt und mit einer id versehen. Die Werte der dropdown-Liste bestimmen als Input-Value in der Callback-Funktion den Output, indem sie einer `fig_data()` als Argument übergeben werden.

Der Einstiegspunkt zur Ausführung des Dashboards ist die Datei `index.py`. Mit dieser Datei wird das Dashboard mittels Flask-Webserver gestartet. Zur Vermeidung zirkulärer Importe ist die Dash-Instanz in der separaten `app.py` definiert [vgl. Plo21f]. Die `index.py` definiert zudem das Layout des gesamten Dashboards. Hier wird auch die Anzahl der

Tabs festgelegt. Die Decorator-Callback-Funktion `@app.callback()` steuert die Auswahl der Tabs und ruft die jeweiligen Tabs auf.

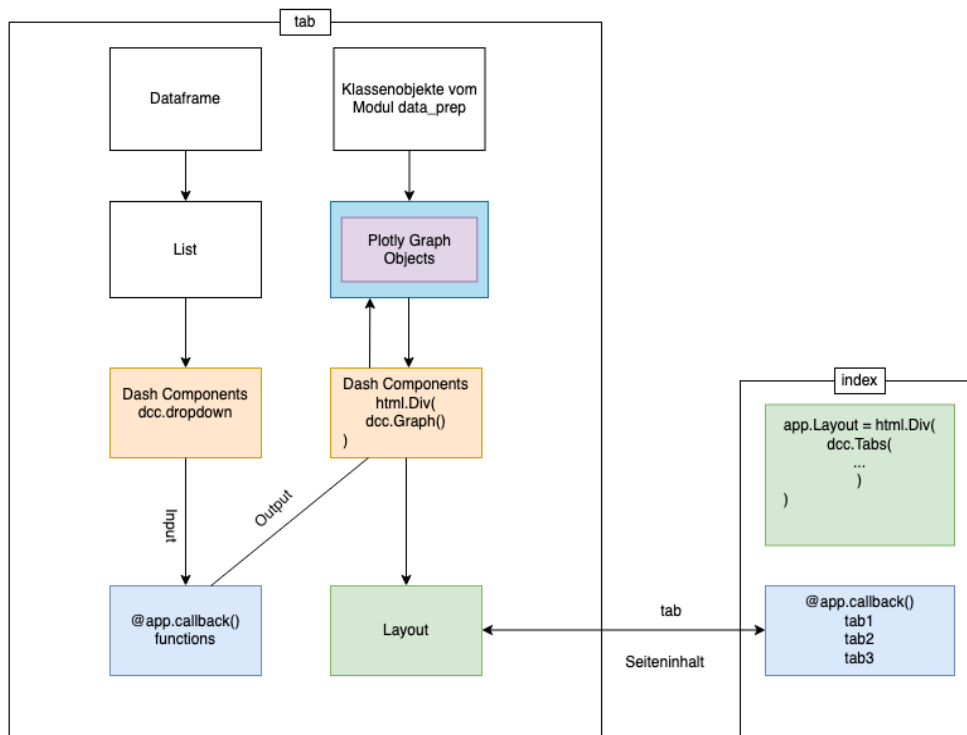


Abbildung 5.10: Ablauf Tab mit index

5.2 FUNKTIONSWEISE

5.2.1 DATEN-IMPORT

5.2.2 DASHBOARD

Dashboard Das Layout des Dashboards besteht aus drei Tabs. Diese wurden nach den Bereichen der Bibliothek aufgeteilt. Diese Aufteilung erschien als sinnvoll, da auf grafische Oberfläche nicht überladen werden sollte. Nach dem der Webserver gestartet wurde, landet man im Tab1 und bekommt die dort verankerten Inhalte zu sehen. Tabs sind als Reiter

oben auf der Webseite angesiedelt. Drauf klicken und man gelangt auf den Inhalt des Tabs (Seite)

Fast allen Diagrammen wohnt inne, dass man aus der Legende auswählen kann welche Elemente angezeigt bzw. nicht angezeigt werden sollen. mitunter auch hover elemente, wenn über die einzelnen Balken oder Linien mit der Maus gefahre

Tab1 - Umsatz und Budget

Bild Tab1 Was ist dargestellt:

zwei Diagramme mit Anzeige des Gesamtumsatz nach Jahren und nach Lieferanten -> horizontalestapeltes Balkendiagramm umsatzstärkste Lieferanten (Top 7) Verteilung in Prozent (Kreisdiagramm)

zwei Diagramme mit interaktiver Auswahl eines Lieferanten über Dropdownmenü Anzeige des Gesamtumsatzes nach Jahren für die ausgewählten Lieferanten (Liniendiagramm) Anzeige des Umsatzes pro Monat für das laufende Jahr (Balkendiagramm)

zwei Diagramme mit Anzeige des Gesamtbudgets nach Jahren und nach Kostenstellen Anzeige kostenintensive (Top 7) Kostenstellen

vier Karten mit Zahl des Gesamtumsatzes Zahl des Umsatzes im Jahr Zahl des Umsatzes im laufenden Jahr für den einzelnen Lieferanten (gekoppelt mit Dropdown-Auswahl) Durchschnittlicher Umsatz des Lieferanten (gekoppelt mit Dropdown-Auswahl)

Tab2 - Lesesaal und Ausleihe

Bild Tab2

zwei Diagramme mit Anzeige der Nutzer:innen nach Jahren und nach Service-Zeiten (gruppiertes Balkendiagramm) Anzeige der Nutzer:innen nach Monaten und nach Service-Zeiten für das laufende Jahr.

ein Diagramm mit Anzeige der Ausleihanzahl nach Jahren (horizontales Balkendiagramm)

Tab3 - Bestand

Bild Tab3

5.3 BEWERTUNG

5.3.1 UMGESETZTE ANFORDERUNGEN

Folgende Must-Anforderungen wurden erfüllt. R1, R2, R3, R4, R5, R7 F1, F4, F5, F12 NF5, NF6,

zur Datenlage Ausleihdaten nur auf Zuruf, mtl. Daten bekommen wir erst seit 2021 zu geschickt. kumulative Daten bei Umsatz/Budget fehlen der UmsatzDaten für Datenbanken müsste nochmal händisch nachgetragen werden (ca. 20.000 Euro) Fehlen der Buchservice-Daten (Subito, ...) -> keine Zeit nur Daten die aus dem CBS/LBS stammen fehlen Daten der Online-Zeitschriften

dennoch: gelungen, die vorhandenen Daten soweit zu importieren, dass mit diesen über das Dataframes Manipulationen und Berechnungen durchgeführt werden können. Ebenso ist es gelungen eine grafische Oberfläche für die Diagramme anzubieten. Diese Diagramme sind mit Plotly interaktiv. Es konnte auch ein bisschen interaktivität selber programmiert werden.

Es konnte gezeigt werden, dass das gelingt...

Den größten Teil der Zeit hat die Datenanalyse gekostet (-> ähnlich Data Science Circle). pandas sehr mächtig und aber auch nicht trivial ist. Ziemlich gutes Tool für die Datenanalyse darstellt.

6 SCHLUSS

6.1 SOLL-IST-VERGLEICH (STAND DER UMSETZUNG)

6.2 LESSONS LEARNED

6.3 WELCHE THEMEN WURDEN NICHT BEARBEITET

Counter-Statistiken

6.4 WELCHE THEMEN SIND IM ANSCHLUSS DENKBAR

bash script für den import der Dateien Daten aus weiteren Datenquellen Integration von Counterstatistiken Integration von RVK-Daten nicht auf der Ebene der Systematikstelle , sondern auf der Ebene der Hauptgruppen -> da Benennung auf Ebene der Systematikstelle zum Teil zu generisch ist. -> sich darein denken erfordert nochmal ein bisschen mehr Zeit. Datenbank-Anbindung Es wurden als Datengrundlage weiterhin nur die Daten berücksichtigt, die für die Anwendungsfälle benötigt werden. Außen vor blieb zum Beispiel die Integration der COP 5-Statistiken.

Refactoring -> wenn System produktiv gehen soll. Zerklopfen der Module -> data_prep falls noch mehr Methoden dazukommen riesige Datei Auslagerung der Chhild classes in eigene Dateien -> flat is always better Wurde nicht alles im OO-Style umgesetzt.

TABELLENVERZEICHNIS

3.1	Informationsdienstleistungen nach Basisfunktionen der Spezialbibliothek	25
3.2	Liste der Dienstleistungsbereiche zu denen statistische Daten erhoben werden	26
4.1	Rahmenbedingungen	30
4.2	Funktionale Anforderungen - ETL-Prozess und Datenspeicherung . .	31
4.3	Funktionale Anforderungen - Datenanalyse	31
4.4	Funktionale Anforderungen - Datenpräsentation und Standardbericht	32
4.5	Nicht-funktionale Anforderungen	33
4.6	Anwendungsfall 1 - Ausleihzahlen Bibliotheksbestand	34
4.7	Anwendungsfall 2 - Ausleihzahlen bibliotheksinterne Lieferdienste . .	35
4.8	Anwendungsfall 3 - Lesesaalnutzung	36
4.9	Anwendungsfall 4 - Neuerwerbungen	37
4.10	Anwendungsfall 5 - Bestandswachstum	38
4.11	Anwendungsfall 6 - Budget- und Umsatzübersicht	39
4.12	Anwendungsfall 7 - Standardbericht	40
5.1	Liste der grundlegenden Programmiersprache und Frameworks	44
5.2	Liste der Teilsysteme	45

ABBILDUNGSVERZEICHNIS

2.1	Statistische Datentypen mit Aussagegehalten und Beispielen	15
2.2	Schichten eines Business-Intelligence-Systems	18
3.1	Anzahl der Titelausleihen nach Identifikationsnummer und Jahr	26
3.2	Monatliche Budget- und Umsatzübersicht	27
5.1	pandas Dataframe und pandas Series	42
5.2	Klassendiagramm - Teilsystem Import	46
5.3	Datenfluss - Teilsystem Import	46
5.4	Monatliche Umsatzübersicht nach Ablauf Teilsystem Import	48
5.5	Klassendiagramm - Teilsystem Datenbearbeitung	49
5.6	Struktur Dash Multi-Page App	52
5.7	Ablauf Tab	53
5.8	Funktionsaufrufe Tab	54
5.9	Ablauf Tab mit Callback	57
5.10	Ablauf Tab mit index	58

QUELLCODEVERZEICHNIS

5.1	Beispiel Methode Exenditures class	50
5.2	Funktion fig_total_expnd() Auszug 1	54
5.3	fig_total_expnd() Auszug 2	55
5.4	html_fig_total_expnd()	55

AKRONYME

BI	Business Intelligence
BIX	Bibliotheksindex
CBS	Zentralsystem
COP 5	Counter 5
COUNTER	Counting Online Usage of NeTworked Electronic Resource
CSV	Comma-separated values
DBS	Deutsche Bibliotheksstatistik
DWH	Data Warehouse
ETL	Extract, Transform, Load
GUI	Graphische Benutzeroberfläche
hbz	Hochschulbibliothekszentrum NRW
HeBis	Hessisches Bibliotheksinformationssystem
KBN	Kompetenznetzwerk für Bibliotheken
KDD	Knowledge Discovery in Databases
KMK	Kultusministerkonferenz
KPI	Key Performance Indicators
LBS	Lokalsystem
mpdl	max-planck-digital-library
MPG	Max-Planck-Gesellschaft
MPI EA	MPI für empirische Ästhetik

Akronyme

NOIR	Nominal-, Ordinal-, Intervall-, Ratio-Systematik
OCLC	Online Computer Library Center
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
OPAC	Online Public Access Catalog
PuRe.MPG	Publikationsrepositorium der Max-Planck-Gesellschaft
RVK	Regensburger Verbundklassifikation
STM	Science, Technology, and Medicine
XML	Extensible Markup Language

LITERATURVERZEICHNIS

- [Alt21] Altair. *Altair: Declarative Visualization in Python — Altair 4.1.0 documentation*. 2021. URL: <https://altair-viz.github.io/index.html> (besucht am 27. 01. 2021).
- [AM17] D. Abts und W. Mülder. *Grundkurs Wirtschaftsinformatik : eine kompakte und praxisorientierte Einführung*. 9., erweiterte und aktualisierte Auflage. Springer, Wiesbaden, 2017. xvi, 758 Seiten.
- [Ban16] C. Bange. „Werkzeuge für analytische Informationssysteme“. In: *Analytische Informationssysteme : Business Intelligence-Technologien und -Anwendungen*. 5., vollständig überarbeitete Auflage. Springer Gabler, Berlin, 2016, S. 114–126.
- [Bec+16] T. Becker, R. Herrmann, V. Sandor, D. Schäfer, und U. Wellisch. *Stochastische Risikomodellierung und statistische Methoden*. Statistik und ihre Anwendungen. Springer, Berlin, 2016. xiv, 375 Seiten.
- [BS04] J. C. Blake und S. P. Schleper. „From data to decisions“. *Library Collections, Acquisitions, & Technical Services* 28:4, 2004, S. 460–464. DOI: [10.1080/14649055.2004.10766018](https://doi.org/10.1080/14649055.2004.10766018). (Besucht am 01. 09. 2020).
- [BS10] J. Bortz und C. Schuster. *Statistik für Human- und Sozialwissenschaftler : mit ... 163 Tabellen*. 7., vollst. überarb. und erw. Aufl. Springer-Lehrbuch. Springer, Berlin, 2010, xvi, 655 Seiten.

- [BS15] M. Block und F. Seeliger. *BibloVis - Ein webbasierter One-Stop Shop zur zentralen Verwaltung und Visualisierung heterogener Nutzungsdaten aus dem Bibliothekskontext*. Type: Konferenzabstract. 104. Deutscher Bibliothekartag (Nürnberg, 2015), 2015.
- [BT20] Baker und Taylor. *Select, Manage and Promote your collection*. 2020. URL: <https://www.collectionhq.com/> (besucht am 28. 09. 2020).
- [Cai16] A. Cairo. *The truthful art : data, charts, and maps for communication*. New Riders, [San Francisco, CA], 2016. xvii, 382 Seiten.
- [Cle11] T. Cleff. *Deskriptive Statistik und moderne Datenanalyse : eine computergestützte Einführung mit Excel, PASW (SPSS) und STATA*. 2., überarbeitete und erweiterte Auflage. Gabler Verlag, Wiesbaden, 2011. xxvi, 227 Seiten.
- [Cou20] Counter. *Abstract | Project Counter*. Abstract | Project Counter. 2020. URL: <https://www.projectcounter.org/code-of-practice-five-sections/foreword/> (besucht am 08. 09. 2020).
- [DDG20] E. Dong, H. Du, und L. Gardner. „An interactive web-based dashboard to track COVID-19 in real time“. *The Lancet Infectious Diseases* 20:5, 2020, S. 533–534. DOI: [10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- [Dea20] Deal. *Projekt Deal*. Projekt DEAL – Bundesweite Lizenzierung von Angeboten großer Wissenschaftsverlage. 2020. URL: <https://www.projekt-deal.de/> (besucht am 08. 09. 2020).
- [Eck11] W. W. Eckerson. *Performance dashboards : measuring, monitoring, and managing your business*. Second edition. John Wiley & Sons, Inc., Hoboken, New Jersey, 2011. xvii, 318 Seiten.
- [Fac21] Faculty. *Dash bootstrap components*. 2021. URL: <https://dash-bootstrap-components.opensource.faculty.ai/> (besucht am 27. 01. 2021).

- [Far11] K. Farkisch. *Data-Warehouse-Systeme kompakt : Aufbau, Architektur, Grundfunktionen*. Springer, Wiesbaden, 2011. xi, 122 Seiten.
- [Few06] S. Few. *Information dashboard design : the effective visual communication of data*. O'Reilly & Associates, Sebastopol, CA, 2006. viii, 211 Seiten.
- [Few09] S. Few. *Now you see it : simple visualization techniques for quantitative analysis*. Analytics Press, Oakland, Calif., 2009. xi, 327 Seiten.
- [Few12] S. Few. *Show me the numbers : designing tables and graphs to enlighten*. Second edition. El Dorado Hills, Calif, 2012. xviii, 351 Seiten.
- [FF16] J. L. Finch und A. R. Flenner. „Using data visualization to examine an academic library collection“. *College and Research Libraries* 77:6, 2016, S. 765–778. DOI: [10.5860/crl.77.6.765](https://doi.org/10.5860/crl.77.6.765). (Besucht am 30. 08. 2020).
- [Gol18] U. Golas. „Statistische Abfragen mit Alma für die Fachreferatsarbeit“. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 5:4, 2018. DOI: [10.5282/o-bib/2018H4S44-57](https://doi.org/10.5282/o-bib/2018H4S44-57). (Besucht am 30. 08. 2020).
- [Gro20] K.-D. Gronwald. *Integrierte Business-Informationssysteme : ganzheitliche, geschäftsprozessorientierte Sicht auf die vernetzte Unternehmensprozesskette ERP, SCM, CRM, BI, Big Data Analytics*. 3., überarbeitete Auflage. Springer Vieweg, Berlin, 2020. XIX, 177 Seiten.
- [HKP12] J. Han, M. Kamber, und J. Pei. *Data Mining : concepts and techniques*. Third edition. Elsevier, Amsterdam, 2012. xxxv, 703 Seiten.
- [HTW18] L. Horne-Popp, E. Tessone, und J. Welker. „If you build it, they will come: creating a library statistics dashboard for decision-making“. In: *Developing In-House digital tools in library spaces*. Hrsg. von L. Costello. IGI Global/-Information Science Reference, Hershey, PA, 2018, S. 177–203.

- [Hug16] M. Hughes. „A long-term study of collection use based on detailed Library of Congress Classification: a statistical tool for collection management decisions“. *Collection Management* 41:3, 2016, S. 152–167. DOI: [10.1080/01462679.2016.1169964](https://doi.org/10.1080/01462679.2016.1169964). (Besucht am 30. 08. 2020).
- [Inm05] W. H. Inmon. *Building the data warehouse*. 4. ed. Wiley, Indianapolis, IN, 2005. xxviii, 543 Seiten.
- [Jil04] C. Jilovsky. „Library Statistics: reflecting yesterday, today and tomorrow“, 2004. URL: https://www.caval.edu.au/assets/files/Research_and_Advocacy/Library_Statistics-reflecting_yesterday_today_and_tomorrow-Northumbria_2005.pdf (besucht am 01. 09. 2020).
- [JM15] J. Johannsen und B. Mittermaier. „Bestands- und Beschaffungsevaluierung“. In: *Praxis Handbuch Bibliotheksmanagement*. Hrsg. von R. Griebel, H. Schäffler, und K. Söllner. Bd. 1. De Gruyter, Berlin, 2015, S. 252–269.
- [Joh14] P. Johnson. *Fundamentals of collection development and management*. Third edition. Ala edition, Chicago, 2014. xiv, 554 Seiten.
- [KBM10] H.-G. Kemper, H. Baars, und W. Mehanna. *Business intelligence - Grundlagen und praktische Anwendungen : eine Einführung in die IT-basierte Managementunterstützung*. 3., überarb. und erw. Aufl. Vieweg + Teubner, Wiesbaden, 2010. ix, 298 Seiten.
- [Kir19] A. Kirk. *Data visualisation : a handbook for data driven design*. 2nd edition. Sage, Los Angeles, 2019. 312 Seiten.
- [KM20] A. Kutlay und C. Murgu. „Shiny Fabric: a lightweight, Open-Source-Tool for visualizing and reporting library relationships“. *Code4Lib* 47, 2020. URL: <https://journal.code4lib.org/articles/14938>.

- [KWC06] J. E. Knievel, H. Wicht, und L. S. Connaway. „Use of Circulation Statistics and Interlibrary Loan Data in Collection Management“. 2006 67:1, 2006, S. 35–49. DOI: [10.5860/crl.67.1.35](https://doi.org/10.5860/crl.67.1.35). (Besucht am 29. 08. 2020).
- [Lai13] M. Laitinen. „Library statistics with confidence: facts from figures with no fear“. *Qualitative and Quantitative Methods in Libraries* 2:4, 2013, S. 459–467. URL: <http://www.qqml-journal.net/index.php/qqml/article/view/122/122>.
- [Lib20a] J. C. K. Library. *JCKL Statistics Dashboard - James C. Kirkpatrick Library - University of Central Missouri*. 2020. URL: <https://library.ucmo.edu/stats/dashboard> (besucht am 28. 09. 2020).
- [Lib20b] E. Libris. *Alma Analytics - Data-Rich & Actionable Decision Support | Ex Libris*. 2020. URL: <https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics/> (besucht am 28. 09. 2020).
- [Lin16] M. Linden. *Geschäftsmodellbasierte Unternehmenssteuerung mit Business-Intelligence-Technologien : Unternehmensmodell - Architekturmodell - Datenmodell*. Springer Gabler, Wiesbaden, 2016. xxiv, 403 Seiten.
- [Lou21] M. Loukides. *Where Programming, Ops, AI, and the Cloud are headed in 2021*. O'Reilly Media. 2021. URL: <https://www.oreilly.com/radar/where-programming-ops-ai-and-the-cloud-are-headed-in-2021/> (besucht am 26. 01. 2021).
- [Lyo10] L. E. Lyons. „Collection evaluation : selecting the right tools and methods for your library“. In: *Library data : empowering practice and persuasion*. Hrsg. von D. Orcutt. Libraries Unlimited, Santa Barbara, Calif., 2010, S. 37–51.

- [MB00] H. Mucksch und W. Behme, Hrsg. *Das Data Warehouse-Konzept : Architektur - Datenmodelle - Anwendungen : mit Erfahrungsberichten*. 4., vollständig überarbeitete und erweiterte Auflage. Gabler, Wiesbaden, 2000. xix, 541 Seiten.
- [Mey18] A. Meyer. „Using R and the Tidyverse to generate library usage reports“. *Code4Lib* 39, 2018. URL: <https://journal.code4lib.org/articles/13282> (besucht am 30. 08. 2020).
- [MH12] E. Morton-Owens und K. Hanson. „Trends at a glance: a management dashboard of library statistics“. *Information Technology and Libraries* 31, 2012. DOI: [10.6017/ital.v31i3.1919](https://doi.org/10.6017/ital.v31i3.1919). (Besucht am 30. 08. 2020).
- [Mic20] Microsoft. *Datenvisualisierung*. 2020. URL: <https://powerbi.microsoft.com/de-de/> (besucht am 28. 09. 2020).
- [ML13] R. M. Müller und H.-J. Lenz. *Business Intelligence*. Springer, Berlin, 2013. xxii, 306 Seiten.
- [Mor15] M. Moravetz-Kuhlmann. „Erwerbungspolitik, Etatplanung und Mittelallokation in wissenschaftlichen Bibliotheken“. In: *Praxis Handbuch Bibliotheksmanagement*. Hrsg. von R. Griebel, H. Schäffler, und K. Söllner. Bd. 1. De Gruyter, Berlin, 2015, S. 161–183.
- [Mur13] S. A. Murphy. „Data visualization and rapid analytics: applying Tableau Desktop to support library decision-making“. *Journal of Web Librarianship* 7:4, 2013, S. 465–476. DOI: [10.1080/19322909.2013.825148](https://doi.org/10.1080/19322909.2013.825148).
- [OCL20] OCLC. *BibControl: Statistik und Reporting für Bibliotheken*. 2020. URL: <https://www.oclc.org/de/bibcontrol.html> (besucht am 28. 09. 2020).
- [Pan21] Pandas. *Pandas - Python Data Analysis Library*. Pandas. 2021. URL: <https://pandas.pydata.org/> (besucht am 26. 01. 2021).

- [Phe12] E. Phetteplace. „Effectively visualizing library data“. *Reference & User Services Quarterly* 52:2, 2012, S. 93–97. DOI: [10.5860/rusq.52n2.93](https://doi.org/10.5860/rusq.52n2.93). (Besucht am 30. 08. 2020).
- [Plo21a] Plotly. *Dash callbacks*. 2021. URL: <https://dash.plotly.com/basic-callbacks> (besucht am 27. 01. 2021).
- [Plo21b] Plotly. *Dash core components*. 2021. URL: <https://dash.plotly.com/dash-core-components> (besucht am 27. 01. 2021).
- [Plo21c] Plotly. *Dash documentation & user guide*. 2021. URL: <https://dash.plotly.com/introduction> (besucht am 27. 01. 2021).
- [Plo21d] Plotly. *Dash html components*. 2021. URL: <https://dash.plotly.com/dash-html-components> (besucht am 27. 01. 2021).
- [Plo21e] Plotly. *Plotly open source graphing libraries*. 2021. URL: <https://plotly.com/graphing-libraries/> (besucht am 27. 01. 2021).
- [Plo21f] Plotly. *URL routing and multiple apps*. 2021. URL: <https://dash.plotly.com/urls> (besucht am 29. 01. 2021).
- [Pyt21a] Python. *6. Modules — Python 3.7.9 documentation*. 2021. URL: <https://docs.python.org/3.7/tutorial/modules.html> (besucht am 26. 01. 2021).
- [Pyt21b] Python. *PyPI : the Python package index*. 2021. URL: <https://pypi.org/> (besucht am 26. 01. 2021).
- [Pyt21c] Python. *Welcome to Python.org : the official home of the Python Programming Language*. Welcome to Python.org. 2021. URL: <https://www.python.org/> (besucht am 26. 01. 2021).

- [Rös+19] H. Rösch, J. Seefeldt, K. Umlauf, und P. Engelbert, Hrsg. *Bibliotheken und Informationsgesellschaft in Deutschland : eine Einführung*. 3., neu konzipierte und aktualisierte Auflage. Harrassowitz Verlag, Wiesbaden, 2019. xiii, 329 Seiten.
- [SAP20] SAP. *Pixel-perfect report creation, data analysis and report distribution*. 2020. URL: <https://www.crystalreports.com/> (besucht am 28. 09. 2020).
- [SB08] R. M. Schmidt und B. Bauer. „Deutsche Bibliotheksstatistik (DBS): Konzept, Umsetzung und Perspektiven für eine umfassende Datenbasis zum Bibliothekswesen in Deutschland: 10 Fragen von Bruno Bauer an Ronald M. Schmidt, Leiter der DBS“. *GMS Medizin - Bibliothek - Informatio* 8:1, 2008, S. 1–7. URL: <http://www.egms.de/en/journals/mbi/2008-8/mbi000102.shtml> (besucht am 08. 09. 2020).
- [Sof20] T. Software. *Software für Business Intelligence und Analytics*. 2020. URL: <https://www.tableau.com/de-de> (besucht am 28. 09. 2020).
- [Spi17] E. T. Spielberg. „Der FachRef-Assistent : personalisiertes, fachspezifisches und transparentes Bestandsmanagement“. Master Thesis. 2017. xiii, 107 Seiten. URL: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:hbz:79pbc-opus-9888> (besucht am 30. 08. 2020).
- [Spr20] SpringShare. *LibInsight - analyze library services and make more informed service decisions*. 2020. URL: <https://springshare.com/libinsight/> (besucht am 28. 09. 2020).
- [Tuf19] E. R. Tufte. *The visual display of quantitative information*. 2. Aufl. Graphics Press, Cheshire, CT, USA, 2019, 197 Seiten.
- [Van21] B. Van de Ven. *Bokeh*. 2021. URL: <https://bokeh.org/> (besucht am 27. 01. 2021).

- [WH13] L. K. Wiegand und B. Humphrey. „Visualizing library statistics using Open Flash Chart 2 and Drupal“. *Code4Lib* 19, 2013. URL: <https://journal.code4lib.org/articles/7812> (besucht am 30.08.2020).
- [WK20] A. Witherley und A. Kirk. *The Chartmaker directory*. 2020. URL: <https://chartmaker.visualisingdata.com/> (besucht am 28.09.2020).

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