

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

Scientific documents often use \LaTeX for typesetting. While numerous packages and templates exist, it makes sense to create a new one. Just because.

ZUSAMMENFASSUNG

Scientific documents often use \LaTeX for typesetting. While numerous packages and templates exist, it makes sense to create a new one. Just because.

INHALTSVERZEICHNIS

1	EINFÜHRUNG	1
1.1	Problemstellung	1
1.2	Ziel der Arbeit	2
1.3	Verwandte Arbeiten	2
1.4	Gliederung der Arbeit	4
2	THEORETISCHE GRUNDLAGEN	5
2.1	Bibliothek und Statistik	5
2.2	Datenvisualisierung	9
2.3	Business-Intelligence-Systeme	13
3	AUSGANGSSITUATION	15
3.1	Bibliothek	15
3.1.1	Allgemeines	15
3.1.2	Organisatorische Einbettung	16
3.1.3	Informationsdienstleistungen	16
3.1.4	Evaluation der Informationsdienstleistungen	17
4	KONZEPTION EINER LÖSUNG	21
4.1	Anforderungsanalyse	21
4.1.1	Ziel	21
4.1.2	Funktionale Anforderungen	21

4.1.3	Nicht funktionale Anforderungen	22
4.1.4	Anwendungsfälle	22
5	DISKUSSION DER UMSETZUNG	23
5.1	Design	23
5.1.1	Systemarchitektur	23
5.1.2	Teilsysteme	23
5.2	Implementierung	23
5.2.1	Umgesetzte Anforderungen	23
5.2.2	Funktionsweise	23
5.3	Bewertung	23
6	SCHLUSS	25
6.1	Soll-Ist-Vergleich (Stand der Umsetzung)	25
6.2	Lessons learned	25
6.3	Welche Themen wurden nicht bearbeitet	25
6.4	Welche Themen sind im Anschluss denkbar	25
	TABELLENVERZEICHNIS	27
	ABBILDUNGSVERZEICHNIS	29
	QUELLCODEVERZEICHNIS	31
	AKRONYME	33
	GLOSSAR	35
	LITERATURVERZEICHNIS	37

1 EINFÜHRUNG

Ausgehend von ökonomischen, informationstechnologischen und marktpolitischen Einschnitten in den vergangenen Jahrzehnten¹, 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 Kennzahlen ist in diesem Zusammenhang größer geworden. Deswegen ist es wichtig, Daten aus bibliothekarischen Servicedienstleistungen und Geschäftsprozessen zu aggregieren, zu erheben und statistisch auszuwerten, um auf Basis der daraus erzielten Erkenntnisse handeln zu können.

1.1 PROBLEMSTELLUNG

Was ist das große Problem?

Budgetplanungen in Bibliotheken immer wichtiger...

Was ist der Trend?

Dashboards in aller Munde

¹ Als Gründe zu nennen wären hier: die Explosion der Zeitschriftenpreise im Bereich der Science, Technology & Medicine (STM), das Aufkommen von E-Publishing und die Konzentration auf wenige Verlage

1.2 ZIEL DER ARBEIT

Schaffung eines Dashboards für Budgetplanung in Bibliotheken Ziel der zu entstehenden Arbeit ist die Entwicklung einer interaktiven Business-Intelligence-Applikation als proof-of-concept, mit der systematisch die relevanten Daten einer hybriden Spezialbibliothek aggregiert, statistisch analysiert und mit geeigneten und modernen Datenvisualisierungstechniken² ausgegeben werden sollen. Vor allem soll sich hier auf automatisierte Prozesse zur Gewinnung der Ergebnisse konzentriert werden.

Mit diesen automatisch angefertigten statistischen Datenanalysen sollen zukünftige Entscheidungen im Bibliotheksmanagement wie Erwerbungspolitik, Budgetplanung und Mittelallokation hinsichtlich der weiteren Entwicklung der Servicedienstleistungen evidenzbasiert und datengetrieben unterstützt werden.

Darüber hinaus soll die Applikation eine Funktion beinhalten, ausgewählte Resultate automatisiert als *factsheet* zu exportieren, um diese als Rechenschaftsbericht gegenüber Stakeholdern der Bibliothek präsentieren zu können.

1.3 VERWANDTE ARBEITEN

Welche Vorarbeiten gibt es schon?

Wo und Wann sind die Vorarbeiten entstanden?

Welche Ziele haben die Vorarbeiten verfolgt?

Auf welche Schwierigkeiten sind sie gestossen?

² Visualisierungen können komplexe Sachverhalte herunterbrechen und so große Datenmengen - im Gegensatz zu großen Tabellen - leicht verständlich darstellen. Im Kontext dieser Arbeit konzentriere ich mich auf Ansätze, die Visualisierungen mittels Visualisierungstechniken algorithmisch aus Daten erzeugen (Informationsvisualisierung, Datenvisualisierung und visuelle Analyse). [[May17](#)]

Es gibt eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf Business-Intelligence-Software basieren. Zu nennen wären *AlmaAnalytics* für das Next-Generation-Library-System *Alma* von *ExLibris*³, *BibControl* von *OCLC*⁴, *CollectionHq* von *Baker & Taylor*⁵ oder *Libinsight* von *SpringShare*⁶. Darüber hinaus gibt es Business-Intelligence-Applikationen, die von Bibliotheken für Reporting, Datenanalyse und Datenvisualisierung adaptiert werden, wie zum Beispiel *Tableau* von der Firma *Tableau Software* oder *Crystal Reports* von *SAP*. Diese Applikationen sind entweder an bestimmte Bibliothekssysteme zurückgebunden, limitiert in ihren Funktionen[[Gol18](#)] oder zu generisch. Überdies wird sowohl von *HeBis* bzw. von der Lokal-Bibliothekssystembetreuung als auch von der *mpdl* keine Applikation in dieser Richtung angeboten. Ebenso ist ungewiss, wann die Ablösung des schon betagten *CBS/LBS* hin zu einem neuen Next-Generation-Library-System im *HeBis-Verbund* stattfinden wird und ob es ein Modul zur statistischen Datenerhebung liefern wird. Ein gutes Beispiel für ein datengetriebenes Unterstützungssystem findet sich in der Literatur 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. Eine Ausnahme bildet die Entwicklung eines Dashboards an der *New York University Health Sciences Libraries*, das versucht, möglichst viele Metriken aus bibliothekarischen Dienstleistungen aufzunehmen.[[MH12](#)] Fast alle Projekte sind an größeren Universitätsbibliotheken mit ganz unterschiedlichen softwaretechnischen Herangehensweisen[[FF16](#); [WH13](#)] und Zielen[[Phe12](#)] entstanden.

³ <https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics> Stand: 26.05.2020

⁴ <https://www.oclc.org/de/bibcontrol.html> Stand: 26.05.2020

⁵ <https://www.collectionhq.com/> Stand: 26.05.2020

⁶ <https://springshare.com/libinsight/> Stand: 26.05.2020

1 Einführung

Dennoch fehlen in der gesichteten Literatur Teile, die sich mit der Budgetierung befassen und Auskunft über Mittelallokation geben.

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.

Anwendung kommen deskriptive Statistik, es geht vielmehr darum Daten zusammen zu tragen, als diese zu explorieren Abgrenzung deskriptive Statistik / explorative Datenanalyse

1.4 GLIEDERUNG DER ARBEIT

Was möchte ich in den folgenden Kapiteln aufzeigen?

Was sind die Kernprobleme mit denen sich die folgenden Kapitel befassen? (Drei bis fünf Sätze pro Kapitel)

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 darauf folgende Abschnitt handelt von Datenvisualisierungen und deren Einsatz für Datenrepräsentationen und Datenpräsentation. Abschließend wird das Modell der Business-Intelligence-Software als Schmelzpunkt der beiden vorangegangenen Kapitel eingeführt.

2.1 BIBLIOTHEK UND STATISTIK

Die Etatplanungen von Bibliotheken richten sich nach deren Informations- und Versorgungsauftrag. Seit Beginn der 1990er Jahre müssen sich Bibliotheken mit den Auswirkungen einer veränderten Medienlandschaft auseinandersetzen. Sie kämpfen mit dem größer werdenden Informationsangebot, den steigenden Preisen auf dem Publikationsmarkt, den zunehmenden Kommerzialisierungstendenzen in der Verlagslandschaft und den 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. Somit geht ein Kaufkraftverlust einher [Vgl. [Mor15](#), 164 ff.]. Diese Entwicklung betrifft nicht nur Universitätsbibliotheken, sondern auch Spezialbibliotheken von Forschungseinrichtungen. Bibliotheken haben Instrumente entwickelt, um den Informationsauftrag trotz dieser Widrigkeiten zu erfüllen. Es entstehen von Bund und Ländern geförderte

Konsortien, um den Kostendruck auf Bibliotheken im Bereich der elektronischen Fachinformationen zu mildern. Neue Geschäftsmodelle werden zur Abfederung der Kosten entwickelt, um Preisnachlässe bei den 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 abschließen [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 Bibliotheksstistik (DBS). Träger der *DBS* sind das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz), das Kompetenznetzwerk für Bibliotheken (KBN), die Kultusministerkonferenz (KMK) sowie den 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]. Neben anderen Servicedienstleistungen bietet die *DBS* Gesamtauswertungen an. Daneben gab es den Bibliotheksindex (BIX), der ursprünglich für die Leistungsmessung in Öffentlichen Bibliotheken konzipiert wurde. 2002 wurde er erweitert auf das Wissenschaftliche Bibliothekssystem. Der *BIX* wurde 2015 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 sammlungs-, nutzungsbezogene und nutzer:innenbezogene Evaluationen zu nennen.[Vgl. Joh14, S. 302] Basieren 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.

Nutzungsbezogene Evaluation umfasst die Lesesaalnutzung, die Ausleihe vor-Ort, die Nutzung des Fernleihservices oder Dokumentenlieferdienste und 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 elektronischer 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 Deakquisition schlecht oder gar nicht genutzter Titel. Ebenso kann die Evaluation von Fernleih- und Dokumentenliefe-

rungen Hinweise auf Bestandslücken liefern [Vgl. [JM15](#), 255 ff.]. Als Konsequenz aus den COUNTER-Statistiken kann die Abbestellung von elektronischen Ressourcen resultieren.

Die nutzer:innenbezogene Evaluation ist auf den Nutzer:innenkreis der Bibliothek und dessen Informationsbedürfnisse zentriert. Die sammlungs- und nutzungsorientierten Evaluationsverfahren basieren auf der Erhebung von quantitativen Daten wie der Bestandsgröße oder der Anzahl von Ausleihen. Nutzer:innenbezogene Evaluation benutzt qualitative Daten, die sie aus Befragungen erhebt.

Die einzelnen Evaluationen vermitteln ein realistisches 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 zufriedenstellendes und korrektes Bild der Situation der Bibliothek zu präsentieren, helfen sorgsam ausgewählte Datenvisualisierungen.

2.2 DATENVISUALISIERUNG

Im Gegensatz zur Inferenzstatistik ist die Aufgabe der Deskriptiven Statistik die Gewinnung von Information aus Daten. Dazu werden verschiedene statistische Verfahren beziehungsweise Methoden angewendet. Datenvisualisierungen sind statistische Verfahren der Deskriptiven und der Explorativen Statistik. [Vgl. Cle11, S. 3; Vgl. Coo21, 7 ff.].

Der Begriff der Datenvisualisierung umschreibt die visuelle Repräsentation und Präsentation von Daten [Vgl. Kir19, 15 ff.]. Er wird in Teilen der Literatur als Oberbegriff für „*Information visualization*“ und „*Scientific Visualization*“ verstanden [Vgl. Few09, S. 11]. Datenvisualisierung grenzt sich in Form und Inhalt von dem Begriff der Infographik ab.

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].

Datenvisualisierungen sollen die Analyse, Exploration und Entdeckung der Daten ermöglichen. Sie sollen das Verständnis der dargestellten Daten erleichtern und sind nicht primär dafür geschaffen, Geschichten über die Informationen zu erzählen [Vgl. Kir19, 20 ff.]. Sie dienen als Werkzeug, um aus den visualisierten Daten Schlussfolgerungen ziehen zu können [Vgl. Cai16, S. 31].

In der Literatur finden sich verschiedene Eigenschaften von Datenvisualisierungen. Datenvisualisierungen sollen auf gründlicher und ernsthafter Forschung basieren. Sie sind funktional, das heißt sie bemühen sich die Daten genau darzustellen. Darüber hinaus sollen Datenvisualisierungen attraktiv, aufschlussreich und erhellend sein [Vgl. Cai16, S. 45]. Ferner sollen sie computerunterstützt und interaktiv sein [Vgl. Few09, S. 12].

Datenvisualisierungen arbeiten mit visuellen Elementen wie Charts, Diagrammen, Tabellen und Karten. Mit dem Einsatz dieser visuellen Elemente sollen Muster, Trends

und Ausnahmen in Daten leichter sichtbar gemacht werden. Datenvisualisierungen setzen einen visuellen Reiz, der schneller vom menschlichen Auge verarbeitet werden kann [Vgl. [Few09](#), S. 32].

Welche Visualisierungen zum Einsatz kommen, hängt auch von der Größe der Datenmenge ab und dem Zweck mit die Datenvisualisierung verfolgt wird. Bei kleineren Datenmengen können die Daten mit dem visuellen Element der Tabelle dargestellt werden. Um Muster, Trends und Ausnahmen aber bei größer werdenden Datenmengen zu erkennen, reichen Tabellen aber nicht mehr aus. Deswegen kommen andere visuelle Elemente wie Diagramme zum Einsatz. Häufig anzutreffende Diagramme sind Balken-, Kuchen oder Linien- und Streudiagramme. Ebenfalls ist für die Auswahl geeigneter Visualisierungen entscheidend, welche Merkmale und Eigenschaften durch die Visualisierung gezeigt werden sollen [Vgl. [Kir19](#), S. 17].

Vor der Wahl der geeigneten statistischen Verfahren ist zu überlegen, welcher Datentyp vorliegt. Datentypen „...define the nature of the values held under each variable and about each item in your dataset.“ [Kir19, S. 99]. Ein Variable (Merkmal) kann qualitativ oder quantitativ sein.

Qualitative Datentypen sind nominale und ordinale Merkmale. Diese werden auch als kategoriale Merkmale bezeichnet. Die Merkmalsausprägungen kategorialer Merkmale können die Werte von einzelnen Zeichen und aber auch von Zahlen annehmen.

Quantitative Datentypen sind metrische Merkmale. Metrische Merkmale können nur numerische Werte annehmen.

Darüber hinaus können die Merkmale diskret oder stetig sein. Ein diskretes Merkmal kann auf Basis der natürlichen Zahlen abzählbar viele Merkmalsausprägungen (values) annehmen. Die Bestandsgröße einer Bibliothek ist ein diskretes Merkmal.

Im Gegensatz dazu können die Merkmalsausprägungen eines stetigen Merkmals jeden beliebigen Wert annehmen. So ist die Raumtemperatur ein stetiges Merkmal.

In der Literatur werden die metrischen Datentypen und die metrische Skala häufig nochmals unterschieden in Interval, Ratio-Datentypen [Vgl. Kir19, 100 f.] bzw. Interval- und Verhältnisskalen [Vgl. BS10, 13 ff.].

Diese Überlegungen bestimmen die Verwendung der visuellen Elemente.

2 Theoretische Grundlagen

Mit dem Siegeszug des Computers in den 1980/90er Jahren sind ...

leicht verschiedene Begriffsdefinitionen Vielzahl von Begriffen

Zu welchem Zweck

Eigenschaften

Wie Datenvisualisierungen gestaltet werden sollen - simpel

Grundlage - Daten - quantitativ / qualitativ Warum Datenvisualisierung wichtig ist?

Was erzählen Datenvisualisierungen mehr als Zahlenkolonnen?

Welche Datenvisualisierungen gibt es?

Wo kommen Datenvisualisierungen zum Einsatz?

quatsch Um sich einen Überblick zu verschaffen welche Daten in einem Unternehmen anfallens, kann man diese auf ein Dashboard sammeln. Dashboards sind Instanzen von Business-Intelligence-Software.

2.3 BUSINESS-INTELLIGENCE-SYSTEME

Was sind Business-Intelligence-Lösungen?

Es gibt eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf Business-Intelligence-Software basieren. Zu nennen wären *AlmaAnalytics* für das Next-Generation-Library-System *Alma* von *ExLibris*¹, *BibControl* von *OCLC*², *CollectionHq* von *Baker & Taylor*³ oder *Libinsight* von *SpringShare*⁴. Darüber hinaus gibt es Business-Intelligence-Applikationen, die von Bibliotheken für Reporting, Datenanalyse und Datenvisualisierung adaptiert werden, wie zum Beispiel *Tableau* von der Firma *Tableau Software* oder *Crystal Reports* von *SAP*. Diese Applikationen sind entweder an bestimmte Bibliothekssysteme zurückgebunden, limitiert in ihren Funktionen[Gol18] oder zu generisch. Überdies wird sowohl von *HeBis* bzw. von der Lokal-Bibliothekssystembetreuung als auch von der *mpdl* keine Applikation in dieser Richtung angeboten. Ebenso ist ungewiss, wann die Ablösung des schon betagten *CBS/LBS* hin zu einem neuen Next-Generation-Library-System im *HeBis-Verbund* stattfinden wird und ob es ein Modul zur statistischen Datenerhebung liefern wird.

¹ <https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics>

Stand: 26.05.2020

² <https://www.oclc.org/de/bibcontrol.html> Stand: 26.05.2020

³ <https://www.collectionhq.com/> Stand: 26.05.2020

⁴ <https://springshare.com/libinsight/> Stand: 26.05.2020

3 AUSGANGSSITUATION

Im folgenden Kapitel wird die wissenschaftliche Spezialbibliothek des *Max-Planck-Institut für empirische Ästhetik* (MPI EA) 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. Dies soll die Basis für die Konzeption und Entwicklung eines datengetriebenen Unterstützungssystems.

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 aus den 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 audiovisuellen Materialien. An Bestand umfasst die Bibliothek zirka 11.000 Bücher, 30 laufende

3 Ausgangssituation

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* (mpdl) unterstützt. Deren Portfolio umfasst vorrangig die zentrale Lizenzierung von relevanten elektronischen Informationsressourcen, die Bereitstellung von Softwarelösungen, das Betreiben des Publikationsrepositoriums *PuRe.MPG* der *Max-Planck-Gesellschaft* (MPG) sowie das Vorantreiben von Open-Access.

Darüber hinaus ist die Spezialbibliothek Teil des *hessischen Bibliotheksverbundes* (He-Bis). Seit Ende 2014 finden die Geschäftsprozesse der Katalogisierung und der Erwerbung im *Zentralsystem* (CBS) und im Lokalsystem *Lokalsystem* (LBS) vom *Online Computer Library Center* (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-Leistung werden der Bibliothek besondere Funktionalitäten für das *Zentralsystem* und Statistiken aus dem *LBS* bereitgestellt.

3.1.3 INFORMATIONSDIENSTLEISTUNGEN

Das Bibliotheks-Team des *MPI EA* ist verantwortlich für den Ablauf und Organisation der bibliothekarischen Informationsdienstleistungen, die der Informationsversorgung dienen. 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 des nutzungsorientierten

Bestandaufbaus ebenfalls eine planmäßige Bestandsentwicklung betrieben. Das Erwerbsprofil 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 Erwerbsprofil 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 (Regensburger Verbundklassifikation (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.

Nutzergruppen

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 Quelle aus der sie 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-	-	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 Dienstleistungsbereichen 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 Nutzungsstatistiken 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.

Monatlich bekommt die Bibliothek Budget- und Umsatzübersichten der Kostenstellen und der Lieferanten zugeschickt. Die Kostenstellen bilden die einzelnen Abteilungen und Forschungsgruppen des Institutes ab. Bearbeitet werden nur die Umsatzübersichten der Lieferanten, um die Umsatzverteilung für die Medien zu steuern. Die Ausgaben für die lokal lizenzierten Datenbanken fehlen in der Aufstellung der Ausgaben.

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-Range des Institutes auf elektronische Ressourcen, die konsortial durch

die MPG lizenziert wurden. Darunter fallen ebooks der Verlage Springer, Wiley oder De Gruyter.

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

4.1 ANFORDERUNGSANALYSE

4.1.1 ZIEL

Als Ziel der nachfolgenden Anforderungen ist eine Priorisierung festzulegen nach dem MoSCo-Prinzip.

4.1.2 FUNKTIONALE ANFORDERUNGEN

Was sind funktionalen Anforderungen?

Speicherort

Wie sollen die Daten importiert werden?

Von wo sollen die Daten importiert werden?

Wie sollen die Daten gespeichert werden?

Wo sollen die Daten gespeichert werden?

Sollen Backups der importierten Daten gemacht werden?

Soll es eine log-Datei geben?

Antwort: zentraler Platz

4 Konzeption einer Lösung

Auswertung der Daten

Welche Daten sollen ausgewertet werden?

Visualisierung der Daten

Welche Visualisierungen sind für die Daten sinnvoll?

Welche Visualisierungen sollen zum Einsatz kommen?

Welche Annotationen sollen zur Anwendung kommen?

Welche Farben sollen zur Anwendung kommen?

Interaktivität

Soll es die Möglichkeit geben aus den Visualisierungen auszuwählen?

Soll es die Filterung der Daten zur Darstellung als Möglichkeit der Interaktivität geben?

Welche Möglichkeiten der Interaktivität soll es geben (Filterung, Highlighting, Animation)

4.1.3 NICHT FUNKTIONALE ANFORDERUNGEN

Was sind nicht-funktionale Anforderungen?

4.1.4 ANWENDUNGSFÄLLE

Was sind Anwendungsfälle (welche Daten aus den bibliothekarischen GG)?

5 DISKUSSION DER UMSETZUNG

5.1 DESIGN

5.1.1 SYSTEMARCHITEKTUR

```
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt
```

Quellcode 5.1: Python example

5.1.2 TEILSYSTEME

5.2 IMPLEMENTIERUNG

5.2.1 UMGESETZTE ANFORDERUNGEN

5.2.2 FUNKTIONSWEISE

5.3 BEWERTUNG

6 SCHLUSS

6.1 SOLL-IST-VERGLEICH (STAND DER UMSETZUNG)

6.2 LESSONS LEARNED

6.3 WELCHE THEMEN WURDEN NICHT BEARBEITET

6.4 WELCHE THEMEN SIND IM ANSCHLUSS DENKBAR

TABELLENVERZEICHNIS

3.1	Informationsdienstleistungen nach Basisfunktionen der Spezialbibliothek	17
3.2	Liste der Dienstleistungsbereichen zu denen statistische Daten erhoben werden	18

ABBILDUNGSVERZEICHNIS

QUELLCODEVERZEICHNIS

5.1	Python example	23
-----	--------------------------	----

AKRONYME

BLX	Bibliotheksindex
CBS	Zentralsystem
COP 5	Counter 5
COUNTER	Counting Online Usage of NeTworked Electronic Resource
DBS	Deutsche Bibliotheksstatistik
hbz	Hochschulbibliothekszentrum NRW
HeBis	Hessisches Bibliotheksinformationssystem
KBN	Kompetenznetzwerk für Bibliotheken
KMK	Kultusministerkonferenz
LBS	Lokalsystem
mpdl	max-planck-digital-library
MPG	Max-Planck-Gesellschaft
MPI EA	MPI für empirische Ästhetik
OCLC	Online Computer Library Center
OPAC	Online Public Access Catalog
PuRe.MPG	Publikationsrepositorium der Max-Planck-Gesellschaft
RDA	Resource Description Access
STM	Science, Technology, and Medicine

GLOSSAR

Business Intelligence	Begriff der Wirtschaftsinformatik, der Verfahren und Prozesse zur systematischen Analyse des eigenen Unternehmens bezeichnet.
Business Intelligence Software	Begriff der Wirtschaftsinformatik, der Verfahren und Prozesse zur systematischen Analyse des eigenen Unternehmens bezeichnet.
\LaTeX	A document preparation system
\mathbb{R}	The set of real numbers

LITERATURVERZEICHNIS

- [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.
- [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.
- [Coo21] F. L. Coolidge. *Statistics : a gentle introduction*. Fourth edition. SAGE, Los Angeles, 2021. xxiv, 512 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).
- [Dea20] Deal. *Projekt Deal*. Projekt DEAL – Bundesweite Lizenzierung von Angeboten großer Wissenschaftsverlage. 2020. URL: <https://www.projekt-deal.de/> (besucht am 08. 09. 2020).

- [Few09] S. Few. *Now you see it : simple visualization techniques for quantitative analysis*. Analytics Press, Oakland, Calif., 2009. xi, 327 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).
- [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).
- [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.
- [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>.
- [May17] Y. Mayank. „Infographics: tools for designing, visualizing data and storytelling in libraries“. *Library Hi Tech News* 34, 2017, S. 6–9. DOI: [10.1108/LHTN-01-2017-0004](https://doi.org/10.1108/LHTN-01-2017-0004). (Besucht am 30. 08. 2020).
- [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).
- [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.
- [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).
- [Rös+19] H. Rösch, J. Seefeldt, K. Umlauf, und P. Engelbert, Hrsg. *Bibliotheken und Informationsgesellschaft in Deutschland : eine Einführung*. 3., neu konzipier-

- te und aktualisierte Auflage. Harrassowitz Verlag, Wiesbaden, 2019. XIII, 329 Seiten.
- [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).
- [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).
- [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).

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