

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	3
1.3	Verwandte Arbeiten	5
1.4	Gliederung der Arbeit	6
2	THEORETISCHE GRUNDLAGEN	7
2.1	Bibliothek und Statistik	7
2.2	Datenvisualisierung	11
2.3	Business-Intelligence-Systeme	15
3	AUSGANGSSITUATION	21
3.1	Bibliothek	21
3.1.1	Allgemeines	21
3.1.2	Organisatorische Einbettung	22
3.1.3	Informationsdienstleistungen	22
3.1.4	Evaluation der Informationsdienstleistungen	23
4	KONZEPTION EINER LÖSUNG	27
4.1	Anforderungsanalyse	27
4.1.1	Ziel	27
4.1.2	Funktionale Anforderungen	27

4.1.3	Nicht funktionale Anforderungen	28
4.1.4	Anwendungsfälle	28
5	DISKUSSION DER UMSETZUNG	29
5.1	Design	29
5.1.1	Systemarchitektur	29
5.1.2	Teilsysteme	29
5.2	Implementierung	29
5.2.1	Umgesetzte Anforderungen	29
5.2.2	Funktionsweise	29
5.3	Bewertung	29
6	SCHLUSS	31
6.1	Soll-Ist-Vergleich (Stand der Umsetzung)	31
6.2	Lessons learned	31
6.3	Welche Themen wurden nicht bearbeitet	31
6.4	Welche Themen sind im Anschluss denkbar	31
	TABELLENVERZEICHNIS	33
	ABBILDUNGSVERZEICHNIS	35
	QUELLCODEVERZEICHNIS	37
	AKRONYME	39
	GLOSSAR	41
	LITERATURVERZEICHNIS	43

1 EINFÜHRUNG

Als Ende des Jahres 2019 der Ausbruch der Covid-19-Pandemie begann, entwickelte die Johns Hopkins University zu Beginn des darauf folgenden Jahres ein Dashboard als Antwort auf den anhaltenden Notstand im Bereich der öffentlichen Gesundheit. Diese visualisiert seit je die gemeldeten Fälle weltweit. Es wurde entwickelt, um Forschern, Gesundheitsbehörden und der breiten Öffentlichkeit ein benutzerfreundliches Instrument an die Hand zu geben, mit dem sich der Ausbruch verfolgen lässt. Zu dessen Datenquellen gehören unter anderem die Informationen der Weltgesundheitsorganisation, staatliche und nationale Gesundheitsämter. Diese wurden aggregiert und verdichtet. So visualisiert das Dashboard mit einer Landkarte und Punkten den Ausbruch. Dazu gibt es Zahlen der bestätigten COVID-19-Fälle, Todesfälle und Genesungen für alle betroffenen Länder[Vgl. [DDG20](#), S. 533] Das faszinierende an dem Dashboard ist, das es gelungen ist, alle relevanten Zahlen auf einer Seite darzustellen und somit der interessierten Öffentlichkeit schnell einen Überblick zu verschaffen.

1.1 PROBLEMSTELLUNG

Ausgehend von ökonomischen, informationstechnologischen und marktpolitischen Einschnitten 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 Kennzahlen ist in diesem Zusammenhang größer geworden.

1 Einführung

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. Transparenz in den Daten sorgt für einen besseren Einsatz des Budgets, bessere Grundlage für die Verhandlungen mit Stakeholdern zweckmäßig alle relevanten Statistiken an einem Ort zu haben und nicht verstreut rumliegen diese dann mit geeigneten Methoden und Verfahren auszuwerten und langfristig und kontinuierlich auszuwerten übersichtlich den stakeholder zu präsentieren. kann geschehen mit herkömmlichen Anwendungen wie Tabellenkalkulationsprogrammen oder mit anderen BI-Systemen...

Es gibt bereits eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf Business-Intelligence-Systemen 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*, *Crystal Reports* von *SAP* oder *Microsoft BI*. 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.

Die Notwendigkeit für diese Applikation ist durch das Fehlen eines zentralen Nachweisortes für bibliothekarische Statistiken in der Bibliothek gegeben. Da die Bibliothek

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

zudem verschiedene Recherche-Systeme den Wissenschaftler:innen anbietet, wäre eine Engführung der statistischen Datenerhebung auf eine Plattform begrüßenswert. 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 *Max-Planck-Institutes für empirische Ästhetik* 2014 gegründet wurde, tritt sie ein in eine Phase, in der ab dem Jahr 2021 Budgetplanungen eine größere Rolle spielen werden.

Ein System jenseits von Excel zu entwickeln ist heute einfacher geworden, da es einerseits einen Markt für solche Anwendungen, der im Schatten von DataScience wächst. So muss man kein ausgewachsener FullStack-Entwickler, um eine Anwendung zu programmieren, sondern man kann zurück greifen auf ausgereifere und mächtige Frameworks wenig aufwendig Daten statistisch auszuwerten, eher Problem, dass es so viele heterogene Daten aus verschiedenen Datenquellen gibt die aufbereitet werden müssen. Dashboards gibt es zwar schon lange, sind aber jetzt auch einfacher geworden zu gestalten. Ad-hoc Realtime Data - Weg bewegen von schwierig aufsetzen DWH hinzu Data Lakes, die Daten auswerten, wenn sie benötigt werden

1.2 ZIEL DER ARBEIT

Das Ziel der Arbeit ist die Schaffung eines Dashboards für Budgetplanung in Bibliotheken. In Anlehnung an Business Intelligence (BI)-Systeme soll ein System als proof-of-concept entstehen, mit dem systematisch die relevanten Daten einer hybriden Spezialbibliothek aggregiert, statistisch analysiert und mit geeigneten und modernen Datenvisualisierungen ausgegeben werden sollen. Um künftigen Anforderungen gewachsen zu sein, soll sie

1 Einführung

modulbasiert programmiert werden und dadurch leicht erweiterbar und eventuell von anderen Bibliotheken nachnutzbar sein.

Warum genau dieses Problem?

Ist Ihr Beitrag völlig neu, oder nur ein Baustein?

Ist Ihr Problem schwer zu lösen oder „straight forward“?

eher schwieriger zu lösen, da generischer Ansatz gewählt werden soll -> abstrahieren ein bisschen von konkreter Implementierung Eher Forschung oder eher Anwendung?

eher anwendung Wenn Sie ein System bauen...

Welche Anfragen / Aufgaben wollen Sie beantworten / lösen können?

Entwicklung eines Workflows für die Aktualisierung der Daten... Framework für template design für Daten, die eingespeist werden sollen Welche Kernfunktionalität soll Ihr System haben?

Automatisierte Prozesse bei der Auswertung mit statistischen Verfahren, Import der Daten soll fast vollständig automatisiert sein Interaktivität -> Multidimensionalität, Eingrenzung der Zeiträume, Auswählen welche Auswertungen nach Medienart ... Domainwissen Auswahl aus mehreren Visualisierungen Bereitstellung der wichtigsten KPI's in einem PDF

Was ist ein typischer (Bedienungs-) Prozess für Ihr System?

1) monatliche Budget und Umsatzzahlen kommen als email, werden automatisch in csv umgewandelt -> werden an große csv / datenbank geschrieben mit der veränderten Datenlage entstehen Veränderungen in den Zahlen -> in Visualisierungen 2) Abfrage des Systems nach Neuerwerbungen mit Systemstellen in der RVK, vierteljährlich, halbjährlich, jährlich -> Kontrolle wie und in welchen Systematikgruppen der Bestand wächst -> RVK-Systematik-Tabelle Wer nutzt Ihr System, Ihren Algorithmus?

Nutzen sollen das System Bibliotheksmitarbeiter:innen und Bibliotheksleitung, auch zum Einspeisen der daten -> Fallback, Fehler abfangen beim Import Wodurch ist dieses Nutzungsverhalten gekennzeichnet?

wenig bis keine Kenntnisse -> soll leicht nutzbar sein ohne große Vorkenntnisse durch automatisierte Prozesse zur Gewinnung der Ergebnisse konzentriert werden. nur Datei in Ordner schieben Import/ Export soll automatisch geschehen.

Anleihen von BI-Systemen in der Architektur

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

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[Phel2] entstanden.

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

Im [Kapitel 2](#) werden die theoretischen Grundlagen für die folgenden Kapitel gelegt. Das Kapitel befasst sich mit den Themen Bibliothek und Statistik, Datenvisualisierung und Business-Intelligence-Systemen. 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](#) 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 [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 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. So entstehen seit Mitte der 1990er Jahre 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.

Bibliotheksstistik 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 Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hzbz), 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 Dokumentenlieferungen 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 deren 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

Datenvisualisierungen sind wirkmächtig. Sie stellen einen Weg da, 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 Datenpunkten 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. Da sich die explorativen Verfahren besonders für große Datenmengen eignen, fand mit der rasanten Entwicklung der Informationstechnologien eine starke Verbreitung dieser Verfahren in den vergangenen Jahren statt. Die Datenvisualisierung hat sich aus den explorativen Verfahren zu einem eigenständigen Fachgebiet der Statistik beziehungsweise der Informatik entwickelt [vgl. [Bec+16](#), 28 f.]. Dies belegt auch eine Vielzahl von Fachliteratur, die in den letzten Jahren veröffentlicht wurden.

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 Infographiken⁵ 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, dass 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

⁵ 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].

Daten gezeigt werden soll. Für die Auswahl der Diagramme ist es wichtig zu bestimmen, ob es sich zum Beispiel um einen Kategorienvergleich, eine Zeitreihe, eine Rangfolge, eine relative Häufigkeit oder eine Korrelation handelt [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 den möglichen Aussagegehalten und Beispielen.⁶

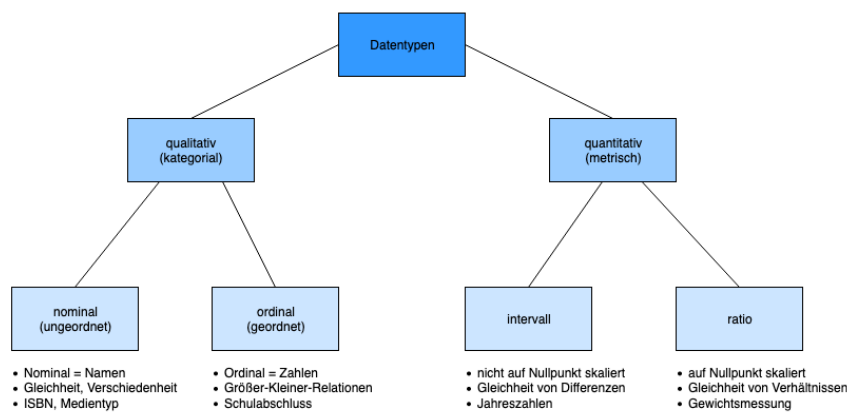


Abbildung 2.1: Statistische Datentypen mit Aussagegehalten und Beispielen

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

⁶ 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].

die Merkmalsausprägungen eines stetigen Merkmals jeden beliebigen Wert annehmen. So ist zum Beispiel die Raumtemperatur ein stetiges Merkmal. Qualitative Merkmale können nur diskret sein, während quantitative Merkmale sowohl diskret als auch stetig sein können. Für kategoriale Merkmale eignen sich Balken- und Kreisdiagramme eher als Liniendiagramme, da kategoriale Daten nur diskret sein können.

Der Einsatz von Datenvisualisierungen wird ebenfalls 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.

Daten können mit Softwareprogrammen wie Microsoft Excel oder anderen Tabellenkalkulationsprogrammen visualisiert werden. In Data-Science-Projekten können ebenfalls verschiedene Frameworks zum Einsatz. 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 guten Überblick erhält man auf der Webseite „The Chart maker“⁷. Eine mächtige und verbreitete Javascript-Bibliothek zur Datenvisualisierung ist D3.js.

Die dargestellten Daten kommen aus allen Bereichen des Lebens. 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 sind Datenvisualisierung ein Bestandteil eines ganzheitlichen Prozess, der unter dem Begriff Business Intelligence (BI) fungiert. Dieser scheint mittlerweile auch im Bibliothekswesen angekommen zu sein.

⁷ <https://chartmaker.visualisingdata.com/> Stand: 17.09.2020

2.3 BUSINESS-INTELLIGENCE-SYSTEME

Business Intelligence (BI)⁸ bezeichnet allgemein Konzepte und Methoden zur Entscheidungsfindung, die auf erfassten Informationen beruhen. Aus technischer Sicht umfasst *BI* die Komponenten Datenhaltung, Datenintegration, Reporting, Data Mining etc. *BI-Systeme* gehören zu den Management-Support-Systemen, die seit den 1960er Jahren in Unternehmen zum Einsatz kommen. [vgl. Gro20, S. 83]

Der Begriff beschreibt nach Abts und Mülder einen integrierten, unternehmensspezifischen, IT-basierten Gesamtansatz zur Unterstützung betrieblicher Entscheidungen [vgl. AM17, S. 270]. Populär wurde er *BI* in den 1990er Jahren und fiel zusammen mit der Entwicklung einheitlich strukturierter und dauerhaft verfügbarer Datenbanken – sogenannten Data Warehouse (DWH) – in Unternehmen. Grund für diese Entwicklung waren die immer neuen Informationsbedürfnisse auf Managementebene, die schnell befriedigt werden sollten [vgl. AM17, 268 f.]. Die Einrichtung dieser Datenbanken zielt auf eine umfassende Informationsversorgung durch Verdichtung von Unternehmensdaten auf unterschiedlichen Stufen mit detaillierten als auch aggregierten Kennzahlen. Der Zugriff auf diese Informationen sollte jederzeit erfolgen können [vgl. AM17, S. 267].

Ein Hauptmerkmal der *BI* ist nach Linden die Entscheidungsunterstützung der Managementebene. [vgl. Lin16, S. 111]. Ein anderes wichtiges Merkmal ist laut Abts und Mülder die Abgrenzung zu operativen Anwendungssystemen [vgl. AM17, S. 267].

BI-Systeme sind nach einer Referenzarchitektur/modell aufgebaut. Dieses Modell umfasst alle technischen Komponenten, die von der konkreten Implementierung aber abstrahiert werden. Die Referenzarchitektur erfasst auch die Datenflüsse zwischen den verteilten Systemeinheiten und Komponenten [vgl. Lin16, 126 ff.]. In der Fachliteratur wird das Modell oft mit drei Ebenen beschrieben [vgl. Lin16, 126 ff. vgl. KBM10, S. 8].

⁸ Synonym wird auch manchmal der Begriff Analytische Informationssysteme oder Business-Intelligence-Systeme gebraucht

Die Ebenen umfassen die Ebene der Datenquellen und Datenerfassung, die Ebene der Datenbereitstellung und Speicherung sowie der Datenanalyse und -präsentation.

Auf der ersten Ebene werden die Daten mithilfe eines Extract, Transform, Load (ETL)-Prozesses bearbeitet. In einem ersten Schritt werden die Daten aus den operativen Vorsystemen extrahiert. Der anschließende Transformationsprozess wandelt die Daten in ein homogenes Format um. Dabei werden die Daten mit Hilfe zum Teil automatischer Verfahren bereinigt, harmonisiert, aggregiert und angereichert [vgl. [AM17](#), 277 f.]. Schließlich werden die Daten im Ladeprozess in ein Data Warehouse (DWH) geladen [vgl. [Lin16](#), 129 ff.].

Das *DWH* entspricht der zweiten Ebene und ist nach Linden der zentrale Bereich der BI-Referenzarchitektur. Ein Data Ware House ist ein logisch zentralisiertes, unternehmensweites Datenhaltungssystem, das physisch von den operativen Vorsystemen getrennt, eine harmonisierte Datenbasis darstellt. Es bietet eine dauerhafte Speicherung der Daten. Damit ist sichergestellt, dass auf diesen Daten längerfristige Analysen durchgeführt werden können [vgl. [Lin16](#), 135 ff.]. In die Konzeptionierung eines *DWHs* sollten ein Archivierungskonzept für die Daten und Überlegungen zu den Aktualisierungszyklen der Datenbank miteinfließen. Die Aktualisierungszyklen legen fest, in welcher periodischen Abfolge die Daten aktualisiert werden, entweder zu Zeitpunkten der Änderungshäufigkeit der Daten im Vorsystem, in periodischen Zeitabständen oder aber auch in Echtzeit.

Die dritte Ebene beschreibt die Ebene der Datenanalyse - und präsentation. Es gibt verschiedene Auswertungskonzepte, die in der Datenanalyse zur Anwendung kommen. Im Rahmen der *BI-Systeme* wird häufig Online Analytical Processing (OLAP) angewendet. OLAP ist eine Anfragetechnik für die Analyse multidimensionaler Daten. Diese Technik ermöglicht, Daten aus verschiedenen Sichtweisen darzustellen. OLAP greift dabei auf die Daten im DWH zu. Ein anderes Verfahren ist Data Mining. Data Mining ist eine explorative Methode, die verschiedene statistische Verfahren auf einen Datenbestand anwendet, um Muster und Zusammenhänge zu erkennen. Data Mining wird auch als Teilphase des

Prozesses der Wissensentdeckung beschrieben vgl. Linden, 142 f. Neuere Tendenzen in *BI-Systeme* führen weg von der strikten Trennung OLTP und OLAP und zur Auflösung des DWHs hinzu Data Lake -> Speicherung der Rohdaten, bis sie benötigt werden -> führt dazu, dass der ETL-Prozess obsolet wird -> flexibler, kosteneffizienter Trennung zwischen OLTP/OLAP-Systemen aufgehoben. Auf Kosten von Persistenz und dauerhafter Sicherung der Daten. In-Memory Computing Hadoop / MapReduce-Verfahren

Die Datenpräsentation umfasst die Datendarstellung. Diese stellt Möglichkeit dar, die zuvor angewendeten Analyseverfahren und Berichtslösungen strukturiert und übersichtlich aufzubereiten Darstellung hochaggregierte, verdichtete quantitativer und qualitativer Information, die für Anwender rasch erfassbar und interpretierbar sind. Die Daten sind auf die Bedarfe der Anwender ausgerichtet und werden auf unterschiedlichen Aggregationsebenen visualisiert.(154) Dabei spielt die Informationsvisualisierung eine wichtige Rolle. Es gibt verschiedene Ziele von Dashboards. 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](#)]

Staging Area -> ETL Dashboards Bilder BI-Systeme OLAP Data-Mining

BI- Systeme (Data Ware House System) mit ganzheitlichen Ansatz Datenbanken, Datenquellen Datenflüsse zwischen den verteilten Systemeinheiten und Komponenten BI-Referenzarchitektur 1. Ebene der Datenquellen- und Datenerfassung Datenextraktionsphase (S. 129 f.) Ausgangspunkt: operative Vorsysteme (Transaktionssysteme) = interne Datenquellen einer Organisation Vereinheitlichung der heterogenen Daten aus verschiedenen Datenquellen und Zusammenführung in einer Datenbank => ETL-Prozess => zur Herstellung einer inhaltlich konsistenten Datenbasis. Festlegung eines Prozess nach Periodizität oder nach Ereignis => Herausziehen der Daten aus den operativen Vorsystemen Datentransformationsphase (130 f.) inhaltliche und strukturelle Umwandlung der Daten in ein homogenes Format Staging Area Filterung => Auswahl, Bereinigung (syntaktische (Sonderzeichen) und semantische Mängel (fehlende Werte) => mögliche Erkennung mit automatisierten Verfahren) und Speicherung der Datenextrakte, Harmonisierung (132)=>???, Aggregation Anreicherung Datenladephase (133) Ladeprozess der die Daten in das Core Data Warehouse (= Gesamtbild der Unternehmensaktivitäten) lädt (mit evtl. Umweg über andere vorgelagerte Datenbanken)

und weiterzuführende betriebswirtschaftliche Analysen ermöglicht 2- Ebene der Datenhaltung und Bereitstellung (135 ff.) zentraler Bereich der BI-referenzarchitektur Core Data Warehouse (= Gesamtbild der Unternehmensaktivitäten) => Bereitstellung einer konsistenten Datenbasis logisch zentralisiertes, unternehmensweites Datenhaltungssystem, das physisch von den operativen Vorsystemen getrennt, eine harmonisierte Datenbasis darstellt, und weiterzuführende betriebswirtschaftliche Analysen ermöglicht dauerhafte Speicherung der Daten, damit längerfristige Analysen durchgeführt werden können -> Zeitreihenanalysen vor Hintergrund historisierter Datenbestände Archivierungskonzepte müssen vorliegen

Es gibt eine Vielzahl kommerzieller Lösungen für den Bibliotheksbereich, die auf Business-Intelligence-Systemen basieren. Zu nennen wären *AlmaAnalytics* für das Next-

Generation-Library-System *Alma* von *ExLibris*⁹, *BibControl* von *OCLC*¹⁰, *CollectionHq* von *Baker & Taylor*¹¹ oder *Libinsight* von *SpringShare*¹². 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.

⁹ <https://www.exlibrisgroup.com/products/alma-library-services-platform/alma-analytics>

Stand: 26.05.2020

¹⁰ <https://www.oclc.org/de/bibcontrol.html> Stand: 26.05.2020

¹¹ <https://www.collectionhq.com/> Stand: 26.05.2020

¹² <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

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)? ¹³

¹³misto

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	23
3.2	Liste der Dienstleistungsbereichen zu denen statistische Daten erhoben werden	24

ABBILDUNGSVERZEICHNIS

2.1	Statistische Datentypen mit Aussagegehalten und Beispielen	13
-----	--	----

QUELLCODEVERZEICHNIS

5.1	Python example	29
-----	--------------------------	----

AKRONYME

BI	Business Intelligence
BIX	Bibliotheksindex
CBS	Zentralsystem
COP 5	Counter 5
COUNTER	Counting Online Usage of NeTworked Electronic Resource
DBS	Deutsche Bibliotheksstatistik
DWH	Data Warehouse
ETL	Extract, Transform, Load
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
NOIR	Nominal-, Ordinal-, Intervall-, Ratio-Systematik
OCLC	Online Computer Library Center
OLAP	Online Analytical Processing
OPAC	Online Public Access Catalog

Akronyme

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

- [AM17] D. Abts und W. Mülder. *Grundkurs Wirtschaftsinformatik : eine kompakte und praxisorientierte Einführung*. 9., erweiterte und aktualisierte Auflage. Springer, Wiesbaden, 2017. xvi, 758 Seiten.
- [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.
- [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).
- [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.

- [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. Schöffler, 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>.
- [Lin16] M. Linden. *Geschäftsmodellbasierte Unternehmenssteuerung mit Business-Intelligence-Technologien : Unternehmensmodell - Architekturmodell - Datenmodell*. Springer Gabler, Wiesbaden, 2016. xxiv, 403 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).
- [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.
- [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 konzipierte 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).
- [Tuf19] E. R. Tufte. *The visual display of quantitative information*. 2. Aufl. Graphics Press, Cheshire, CT, USA, 2019, 197 Seiten.
- [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