

Das Digitale Lernlabor Cottbus

Einblicke in den Aufbau eines kommunalen Data Warehouses
Endbericht

Forschungsprogramm

„ExWoSt – Experimenteller Wohnungs- und Städtebau“, ein Forschungsprogramm des Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Auftrag des Bundesministeriums für Wohnen, Stadtentwicklung und Bauwesen (BMWSB)

Projektlaufzeit

01.09.2018 bis zum 30.11.2021

Aktenzeichen

10.04.05-17.01001

im Auftrag

des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR)
im Bundesamt für Bauwesen und Raumordnung (BBR)

bearbeitet von

Dr. Gustav Lebhart, Stadt Cottbus/Chósebuz

Peter Schirrgott, Stadt Cottbus/Chósebuz

Eva Schweitzer, Referat RS 5 „Digitale Stadt, Risikovorsorge und Verkehr“

Inhaltsverzeichnis

Kurzfassung	3
1. Planung und Steuerung des Datenmanagements	4
1.1. Strategische Ausrichtung	4
1.1.1. Variante A Team „Statistik und Stadtforschung“	5
1.1.2. Variante B Servicebereich „Datenmanagement und Statistik“	5
1.1.3. Variante C Fachbereich „Data Governance“	5
1.2. Mitarbeiterkompetenzen und Kompetenzprofile	6
1.3. Kompetenzprofile im Überblick	7
2. Datenstrategien und Datenqualitätsmanagement	9
2.1. Leitlinie „Datenplanung“	10
2.1.1. Der Top-Down-Ansatz	11
2.1.2. Der Bottom-Up-Ansatz	12
2.1.3. Der Hybrid-Ansatz am Beispiel von Data Vault	13
2.2. Leitlinie „Datenerhebung“	14
2.2.1. Zielarchitektur	14
2.2.2. Anforderungen an das Datenmanagement	15
2.3. Leitlinie „Datenqualität“	16
2.4. Leitlinie „Datenanalytik“	18
2.5. Leitlinie „Datenvisualisierung“	19
2.5.1. Gestaltungsebenen der Datenvisualisierung	19
2.5.1.1. Auswahl-Matrix: Diagrammarten	20
2.5.1.2. Auswahl-Matrix: visuell-retinale Variablen	21
2.5.2. Diagrammarten	21
2.6. Leitlinie „Datenschutz und Datensicherheit“	27
2.6.1. Rechtsgrundlagen eines kommunalen Informationsportals	27
2.6.2. Informationsportal – Einführungsprozess und erforderliche Maßnahmen	28
2.6.2.1. Antragstellung (bei der für die Beschaffung zuständigen Stelle)	28
2.6.2.2. Konzeptions- und Planungsphase	29
2.6.2.3. Implementierungs- und Testphase des kommunalen Informationsportals	30
2.6.2.4. Begleitung und Überwachung der Wirkbetriebsphase	30
2.6.3. Datensicherheit und Nutzung von Cloud-Diensten	31
3. Systemarchitektur für das Datenmanagement	33
3.1. Bedarf und Anforderung in der kommunalen Verwaltung	33
3.2. Data Warehousing und Datenbanksysteme	34
3.2.1. DWH mit Datavault Builder™	35
3.2.2. DWH mit Qlik Compose	35
3.2.3. Self-Service Analytics mit Qlik Sense	36
3.2.4. Tableau Plattform (DWH light)	36
3.3. Datenanalyse- und Datenvisualisierungstool	37
3.3.1. Datenanalytik	37
3.3.2. Datenvisualisierung	38
3.4. Fazit	39

Kurzfassung

Die kreisfreie Stadt Cottbus/Chósebuz hat im Rahmen der experimentellen Stadtforschung (ExWoSt) des Bundes in einem Modellvorhaben Strategien und Möglichkeiten des Aufbaus eines städtischen integrierten Datenmanagementsystems in Form eines „Urban Data Warehouse“ entwickelt und erprobt.

Die Ausgangssituation: Daten werden in der Verwaltung von unterschiedlichen Organisationseinheiten und innerhalb der vielfältigen städtischen Aufgabenbereiche verwendet. Auf kommunaler Ebene zeigt sich eine heterogene Datenlandschaft, die eine Vielzahl an Herausforderungen bei der Nutzung von Daten mit sich bringt. Sowohl unterschiedliche Datenstrukturen und -formate als auch ungleiche Merkmalsausprägungen von Daten offenbaren, bis zu einem gewissen Ausmaß, ein divergierendes Verständnis bezüglich der Bedeutung von Daten in den jeweiligen Fachbereichen der kommunalen Verwaltung. Gleiche Angaben in verschiedenen Systemen bewirken eine unkontrollierte Redundanz. Die Daten sind widersprüchlich, unvollständig und fehlerhaft. Aufgrund dieser Fakten nimmt der Nutzen für die Stadtplanung ab. Dies wird durch etablierte Datensilos einzelner Fachbereiche und fehlende durchgängige digitale Datenflüsse deutlich.

Die Aufgabe: Gesucht wurde somit ein Ansatz zur Datenintegration, welcher die individuellen Aspekte der kommunalen Verwaltung der Stadt Cottbus/Chósebuz berücksichtigt, und zwar die Schaffung einer einheitlichen, jederzeit aktuellen und verlässlichen Datenquelle, also des Single Point of Truth (SPOT) oder des „einzigen Punktes der Wahrheit“ als Grundlage für datenbasierte Auswertungen, Statistiken, Prognosen und letztlich Entscheidungen auf der Basis vertrauenswürdiger Informationen.

Das Ziel: Angestrebt wird die Konzeptentwicklung, wie ein vernetztes und integriertes System der Data Governance, eine Zusammenfassung von Prozessen, Rollen, Richtlinien, Standards und Kennzahlen, in der Stadtverwaltung Cottbus/Chósebuz implementiert werden kann und somit Fragen der integrierten Stadtentwicklung qua optimierter Datenanalyse gelöst werden können. Zudem sollte untersucht werden, welche Visualisierungs- und Darstellungsformen genutzt werden könnten, um durch eine attraktive Darstellung und Veranschaulichung städtischer Daten und Informationen eine stärkere Beteiligung der Zivilgesellschaft zu erreichen.

Das Vorgehen: In einem mehrstufigen Verfahren wurden sowohl Konzepte für ein verwaltungsinternes Organisations- und Rollenmodell des Datenmanagements (Kapitel 1) als auch Leitlinien zum Umgang mit den kommunalen Datenbeständen (Kapitel 2) entwickelt. Im Kapitel 3 werden zudem praktische Hinweise und Empfehlungen zu unterschiedlichen technischen Systemarchitekturen und den möglichen Datenintegrations- und Business Intelligence-Systemen (BI-Systemen) gegeben.

Die strategische Konzeption eines verwaltungsinternen Organisationsmodells (Kapitel 1) wurde federführend von dem Projektleiter und Chief Information Officer (CIO) der Stadt Cottbus/Chósebuz, Dr. Gustav Lebhart, entwickelt. Als Aufgaben im Rahmen des Urban Data Lab (UDL) wurden im Zeitraum von Juli bis September 2020 eine Ist-Analyse der Datenbestände und Datenflüsse der drei Bereiche Team Statistik, Fachbereich Stadtentwicklung sowie Fachbereich Geoinformation und Liegenschaftskataster durchgeführt und anschließend eine Anforderungsanalyse zu dem Thema „Data Warehousing und Datenintegration“ erstellt. Auf der Basis dieser Ergebnisse wurde eine Marktbetrachtung für Datenintegrationswerkzeuge abgeschlossen und es wurden die Leitfäden „Datenplanung“, „Datenerhebung“ und „Datenqualität“ erarbeitet. Für diese Aufgaben wurde im Verlauf des Projektes ein Subauftrag an die Andreas Krohn & Birk Heinze GbR erteilt. Neben den Fragen zum Aufbau einer Datenintegrationsplattform bestand das Ziel der Stadt Cottbus/Chósebuz auch in der Entwicklung eines städtischen Informationsportals auf der Basis der integrierten Daten. Hierfür wurden die zu beachtenden rechtlichen Grundlagen eingearbeitet (Leitfaden Datenschutz und Datensicherheit), Grundsätze der Datenanalyse beschrieben (Leitfaden Datenanalytik) sowie Formen der möglichen Datenabbildung untersucht (Leitfaden Datenvisualisierung) und zu diesem Zweck geeignete Softwareprodukte (Kapitel 3) auf ihre Eignung für die kommunale Praxis geprüft. Mit der Prozessunterstützung in diesem Aufgabenbereich wurde die Syncwork AG beauftragt.

1. Planung und Steuerung des Datenmanagements

Verwaltungsdaten sind eine unverzichtbare Grundlage (nicht nur) für kommunale Planungsaufgaben, um gegebenenfalls vorausschauende Anforderungen zielgerichtet für Infrastruktureinrichtungen ermitteln zu können. In der vorliegenden Konzeption werden strategische Optionen hinsichtlich des Aufbaus und der Weiterführung eines kommunalen integrierten Datenmanagements vor dem Hintergrund der zu erfüllenden Planungsaufgaben des Fachbereiches Stadtentwicklung der kreisfreien Stadt Cottbus/Chósebus skizziert.

Diese konzeptionellen Überlegungen bilden einen Ordnungsrahmen für die Strukturierung des Cottbuser Datenmanagements. Aus Sicht der Verfassenden sollte sich die konzeptionelle Ausgestaltung einer datengetriebenen Organisationseinheit an drei zentralen Prinzipien orientieren:

- Das bestehende Ressourcenpotenzial ist so weit wie möglich zu nutzen
(Nutzung querschnittsorientierter Fachkompetenzen)
- Der Personalaufwand bei der Datenerhebung und Datenauswertung ist gering zu halten
(Ressourcen optimieren statt Neuaufnahme)
- Ein positives Aufwand-Nutzen-Verhältnis in Bezug auf eine höhere Datenqualität und Serviceleistung ist anzustreben *(Verbesserung nicht um jeden Preis)*

Das Datenmanagement definiert Abläufe und Verantwortlichkeiten, um relevante Daten etwa für die Stadtplanung und Stadtentwicklung sicherzustellen.

1.1. Strategische Ausrichtung

Das kommunale Datenmanagement soll das Angebot an statistischen Informationen sowie offenen Verwaltungsdaten koordinieren. Es hat die Tätigkeiten und Fähigkeiten der einzelnen „Akteure“ in der Stadtverwaltung zu vernetzen und soll Aufgaben, Verantwortlichkeiten und Rollen definieren und klären. Mit einem gebündelten Ansatz können fachbereichsübergreifende Prozesse harmonisiert sowie der Zugang zu vielen Datenquellen für Kundeninformationen und Erfolgsmethoden in Bezug auf die Datenverwaltung sichergestellt werden. Ziel muss es daher sein, dass für verwaltungsrelevante Entscheidungen statistische Daten als Grundlage repräsentativ und vollumfänglich von einer übergeordneten Organisationseinheit bereitgestellt werden. Dies schließt unterschiedliche Versionen benötigter Daten aus. Die Datensicherung ist nur an einem Speicherort erforderlich.

Für die Entwicklung einer leitenden Stelle für die Datenaufbereitung sind die Daten der Stadtverwaltung nach internationalen Standards zu konsolidieren. Dadurch wäre der Zugriff für die Fachbereiche in der kommunalen Verwaltung einfacher und eine Vereinheitlichung bei den Verwaltungs- und Standardisierungsprozessen effizienter zu administrieren. Die Datenverwaltung in den Behörden hat sich in den letzten Jahren von einer primär betrieblichen Funktion zu einer strategischen Arbeit entwickelt, die auf eine übergreifende Geschäftsanalytik, die Business Intelligence (BI) ausgerichtet ist. Im Folgenden wird am Beispiel der Stadt Cottbus/Chósebus modellhaft skizziert, wie die unterschiedlichen Aufgaben eines integrierten Datenmanagements auf die einzelnen Fachbereiche verteilt und entsprechende Organisationsprofile sowie Aufgabenbeschreibungen für verschiedene Organisationseinheiten entwickelt werden können. Maßgeblich wurde dieses Rollenmodell von dem damaligen Projektleiter und CIO der Stadt Cottbus/Chósebus, Dr. Gustav Lebhart, entwickelt.

1.1.1. Variante A | Team „Statistik und Stadtforschung“

Das Team „Statistik und Stadtforschung“ könnte organisatorisch in bereits bestehende Strukturen des Geschäftsbereiches Stadtentwicklung und Bauen integriert werden. Durch eine organisatorische Eingliederung sind Synergieeffekte und operative Akzente der fachübergreifenden Zusammenarbeit mit den Fachbereichen Stadtentwicklung sowie Geoinformation und Liegenschaftskataster zu erwarten. Im Hinblick auf die ordnungsgemäße Durchführung der Verwaltungsaktivität koordiniert und steuert das Team alle internen Anfragen. Dies setzt jedoch voraus, dass das Datenmanagement mit entsprechenden BI-Anwendungen für die Analysetätigkeit integriert wird. Die Begutachtung beziehungsweise Evaluierung von externen Studien und Expertisen sowie die Zuarbeit für andere Fachbereiche im Bereich der Datenanalyse und Prognostik gehören zu den hauptsächlichen Aufgaben.

Neben der zentralen Datenaufbereitung werden statistische Informationen übersichtlich dargestellt und einfach verständlich präsentiert. Es stehen aussagekräftige Berichte sowie eine Vielzahl von kleinräumigen Daten für Cottbus/Chósebus zur Verfügung. Mittels übersichtlicher Tabellen, Zeitreihen und Grafiken können sich Interessierte über die wirtschaftlichen, sozialen und politischen Gegebenheiten für die ausgewählten Themen einen schnellen Überblick verschaffen. Aber auch weiterführende räumliche Analysen (zum Beispiel Verschneidung mit vorhandenen Datengrundlagen aus anderen Fachbereichen) sind vom Team durchzuführen.

1.1.2. Variante B | Servicebereich „Datenmanagement und Statistik“

Der Servicebereich, als zentrale Dienstleistungseinheit einer Stadtverwaltung, positioniert sich als strategische und konzeptionelle Anlaufstelle der Verwaltungsführung sowie der Geschäfts- und Fachbereiche. Dieser wird umfassend bei datengesteuerten Informationsprozessen und statistischen Auswertungen eingesetzt. Alle städtischen Planungsdaten (unter anderem Bevölkerungs-, Bau-, Wirtschafts- und Verkehrsdaten) sind in einem Data Warehouse (DWH) gebündelt und können von allen Fachbereichen abgerufen werden. Der Servicebereich „Datenmanagement“ wäre eine neue fachübergreifende Einrichtung in der kommunalen Verwaltung. Mit der organisatorischen Etablierung eines zentralen Datenmanagements mit klaren Verantwortlichkeiten und dokumentierten Abläufen werden mit der richtigen Technologie die Qualität und die Verwertbarkeit der Daten/Statistik erheblich verbessert.

Mit Blick auf die Planungsaufgaben sind auch technologische Anpassungen und Adaptierungen bei bestehenden Fachverfahren unausweichlich. Eine Erweiterung der DWH-Infrastrukturlandschaft und der Einsatz neuer technologischer Software zur Unterstützung von Analyseverfahren sind anzudenken. Bei der Einführung neuer Fachverfahren ist eine enge und stufenweise Zusammenarbeit mit anderen Fachbereichen verbindlich von beiden Seiten einzuhalten. Die zunehmende Komplexität im Bereich der Datenanalyse erfordert eine angemessene Validierung der erforderlichen Fachkompetenz für das Aufgabengebiet.

1.1.3. Variante C | Fachbereich „Data Governance“

Die Aufgabenstellung des Fachbereiches „Data Governance“ ist sehr heterogen, weitestgehend überschneidungsfrei mit den Aufträgen anderer Fachbereiche und unterscheidet sich somit wesentlich vom Tätigkeitsbereich der amtlichen Statistik. Durch die querschnittsorientierten Themen übernimmt die „Data Governance“ einer Verwaltung auch zusätzliche Aktivitäten, vor allem im Bereich der Datenanalyse, und erstellt Sonderauswertungen und Expertisen zu verschiedenen Fragestellungen aus dem Bereich der amtlichen Statistik. Die Statistiken werden auf Logik und Inhalt geprüft und datentechnisch in ein Datenbanksystem (Data Warehouse) eingefügt. Die Art und Weise der Datenspeicherung und -verwaltung über ein Datenbankmodell ist vom Fachbereich festzulegen. Wie im Servicebereich „Datenmanagement und Statistik“ ist unter Berücksichtigung alternativer Architekturvarianten und Technologien das Programmiergerüst für das Daten- und Analysemanagement in eine Strategie aufzunehmen und zu bewerten.

Der Fachbereich stellt ein breit gefächertes Angebot an Zahlen und Fakten aus der Verwaltung bereit. Hier stehen Berichte, Tabellen, Datenbanken, interaktive Karten sowie Publikationen zu einer Vielzahl von gesellschaftlich interessanten Themen zur Verfügung. Die Auswertungen stehen zum kostenlosen Download bereit und werden über eine eigene Datenplattform der Öffentlichkeit zugänglich gemacht. Darüber hinaus werden Publikationen mit Datenmaterial und Statistiken angeboten. Zudem besteht die Möglichkeit, Auswertungen interaktiv selbst vorzunehmen.

1.2. Mitarbeiterkompetenzen und Kompetenzprofile

Eine der großen Herausforderungen bei der Einführung eines fachbereichsübergreifenden Datenmanagements besteht in der Beantwortung der Frage, wie die Datenvielfalt einer Verwaltung geregelt werden soll. Jeder Fachbereich einer kommunalen Verwaltung verfügt über eine Vielzahl an unterschiedlichen Daten. Jedes Datenbank- und Datenfilesystem ist mit unterschiedlichen Überlegungen bezüglich der Qualität, Sicherheit und Compliance aufgebaut worden.

Die Implementierung von Data Governance-Regelungen kann datenadministrative Prozesse und Risiken reduzieren sowie das Vertrauen in organisatorische Informationen erhöhen, und zwar durch

- die Verringerung von Variationen in der Data-Architektur und Datenmodellierung,
- die Standardisierung der Semantik,
- ein Metadaten-Management, um die Transparenz und Sichtbarkeit vom Konzept bis zur technischen Implementierung und Umsetzung zu gewährleisten, und
- die umfassende Daten-Bedarfsanalyse, um alle prospektiven Datenanforderungen der Organisationseinheiten zu erfassen.

Eine „Data Governance-Unit“ besteht in der Regel aus einem Team, dessen Ziel es ist, klare Funktionen zu etablieren und wo jedes Mitglied

- mit seiner fachspezifischen Kompetenz eine gesonderte Rolle einnimmt,
- Verständnis für die Aufgaben der Rolle hat,
- sich bewusst ist, in welcher Funktion welche „Position“ zu vertreten ist und
- nach kritischer Evaluierung eine Priorisierung von Projekten vorzunehmen hat.

Data Governance Committee (DGC): Ist die Instanz zur Steuerung und Überwachung der Data Governance-Programme und für die Umsetzung der Datenqualitätsstrategie verantwortlich. Es stellt das Bindeglied zwischen den operativen Ebenen und den Fachbereichen dar und ist für die Festlegung von (Metadaten-)Standards zuständig. Das Komitee erarbeitet zudem Vorgaben für Projekte zur Verbesserung der Datenqualität und könnte das Projektportfolio überwachen. Folgende Funktionsträger sollten dem DGC angehören: Chief Information Officer (CIO), Chief Digital Officer (CDO), Koordinatoren für Geographische Informationssysteme (GIS) und Datenschutzbeauftragte.

Data Stewards (DS): Sind Beschäftigte aus dem Fachbereich, die ein bestimmtes thematisches Gebiet oder Datendomänen verwalten. Ein Data Steward erarbeitet Grundsätze für die Datenqualität und -sicherheit, plant

die Anforderungen und koordiniert die Datenlieferung. Das Arbeitsfeld kann die Entwicklung von Geschäftsregeln für Daten sowie die Entwicklung und Implementierung von Datenmodellen umfassen.

Sie bringen Expertise in Vorschläge für Standards und Richtlinien ein.

Data Experts (DE): Data Experts nutzen zwei Quellen, um Potenziale aufzudecken. Sie halten einerseits engen Kontakt zu den Fachbereichen und werden dabei mit den brennenden Themen und Fragestellungen konfrontiert.

Andererseits nutzen sie die Daten und kreativen Analysemethoden, um Auffälligkeiten nachzugehen und das Management oder die Fachabteilungen über mögliche geschäftsrelevante Potenziale zu informieren und Entscheidungsvorlagen zu entwerfen.

Data User (DU): Sie nutzen die Daten unter Einhaltung der festgelegten Grundsätze. DU kennen die Datenquellen, verstehen die verwendeten Fakten und extrahieren diese für die spezifischen Bedürfnisse.

Für die Stadtverwaltung Cottbus/Chósebusz wurde im Rahmen des Forschungsprojektes das Datenmanagement vom Projektleiter und CIO der Stadt Cottbus/Chósebusz mit folgenden Kompetenzprofilen konzipiert.

1.3. Kompetenzprofile im Überblick

Tabelle 1: Überblick über die verschiedenen Kompetenzprofile (Quelle: Dr. Gustav Lebhart)

Data Governance-Rolle	Funktion	Aufgabe	Beschreibung
CIO	<ul style="list-style-type: none"> ■ Steuerung 	<ul style="list-style-type: none"> ■ Überwachung der Umsetzung und Strategie-Einhaltung in Bezug auf alle Data Governance-Themen ■ Bestätigung und Entscheidung über Datenmanagement-Strategie sowie notwendige Budgetierung ■ Entscheidung über fachunabhängige oder fachübergreifende IT-Interoperabilitäts- und IT-Sicherheitsstandards sowie Steuerung von E-Government-Projekten ■ Leitung Projektportfolio-Management 	<ul style="list-style-type: none"> ■ Überwacht die Umsetzung und Strategie-Einhaltung aller Datenqualitätsmanagement-Themen (DQM) ■ Trifft Entscheidungen bezüglich DQM-Strategie und klärt Fragen der notwendigen Budgetierung ■ Priorisiert fachliche Datenbedarfe für die Umsetzung sowie Prüfung und Validierung von KPIs für DQM-Steuerungssystem ■ Für eine ressortübergreifende Innovationsstrategie wird der Dialog zwischen relevanten Stakeholdern aus Wirtschaft, Wissenschaft und Verwaltung sowie Vernetzung mit bestehenden Gremien und Plattformen fortgesetzt
CDO	<ul style="list-style-type: none"> ■ Planung ■ Steuerung 	<ul style="list-style-type: none"> ■ Entwurf einer übergreifenden Digitalstrategie und Überwachung ihrer Umsetzung ■ Erkennen von Trends und neuen Technologien ■ Vorantreiben von Innovationen 	<ul style="list-style-type: none"> ■ Analysiert und erkennt alle neuen Trends und Technologien und beurteilt Dynamik und/oder Beständigkeit von Trends ■ Fokussiert Veränderungen im Verwaltungsumfeld ■ Verantwortlich für die Definition und Ausführung von neuen (digitalen) Modellen und Strategieveränderungen ■ Führt Machbarkeits- sowie Marktanalysen durch und zeigt damit den Fachbereichen Innovationspotenziale auf

Data Governance-Rolle	Funktion	Aufgabe	Beschreibung
Koordinator	<ul style="list-style-type: none"> ■ Koordination ■ Evaluierung ■ Steuerung 	<ul style="list-style-type: none"> ■ Leitung Fachbeirat Data Governance (Data Gov) ■ Erstellung von Expertisen zu Themenbereichen des Datenqualitäts- und Informationsmanagements ■ Geben von Empfehlungen und Entwicklung von Richtlinien für Qualitätsstandards und Datenaufbereitung ■ Erarbeitung von Regelwerk und Datenleitlinien für sachgerechte Datenerhebung und statistische Nutzung 	<ul style="list-style-type: none"> ■ Definiert mit dem Fachbeirat Prozesse und Richtlinien für Datennutzung und -pflege, orientiert sich dabei an strategischen und operativen Zielen der öffentlichen Verwaltung ■ Setzt Entscheidungen und Beschlüsse um, koordiniert und überwacht ■ Unterstützt die Kommunikation von CIO-Entscheidungen gegenüber betroffenen Interessengruppen und Stakeholdern ■ Initiiert Vorhaben zum Datenqualitätsmanagement, koordiniert und prüft Datenlieferungen ■ Data Stewards aus den jeweiligen Geschäftsbereichen berichten an Governance-Koordinator
Data Steward	<ul style="list-style-type: none"> ■ Koordination ■ Standardisierung 	<ul style="list-style-type: none"> ■ Mitwirkung an Datenqualitätsstrategie und Konzipierung von Datenqualitätsprinzipien (Plausibilitätsprüfung) ■ Evaluierung von Vorhaben zu Datenqualitätsmanagement und Datenbereitstellung ■ Vertretung in Gremien ■ Kann für den Fachbeirat Data Governance nominiert sein 	<ul style="list-style-type: none"> ■ Vorschläge für Standards und Richtlinien einbringen ■ Als Fachbeiratsmitglied koordiniert ein Data Steward andere Data Stewards und fungiert als deren Sprachrohr im Fachbeirat ■ kann Grundsätze für Datenqualität und -sicherheit detailliert erarbeiten, plant Anforderungen und koordiniert Datenlieferung ■ Kann die Entwicklung von Geschäftsregeln für Daten sowie die Entwicklung und Implementierung von fachlichen Datenmodellen umfassen ■ Beschäftigt sich mit Fragen zur fachlichen Datenarchitektur und formuliert fachliche Anforderungen in Fachkonzepten als Vorgabe ■ Liefert standardisierte Datendefinitionen und -formate, dokumentiert Quellsysteme für Datenobjekte sowie Datenflüsse zwischen Informationssystemen und stimmt diese mit Data Experts ab ■ Fachliche und technische Data Experts aus den Fachbereichen berichten an Data Steward
Data Expert	<ul style="list-style-type: none"> ■ Evaluierung ■ Standardisierung 	<ul style="list-style-type: none"> ■ Unterstützung der Data User beim Verstehen und Nutzen der fachlichen Datenqualitätsanforderungen ■ Begleitung der Data User in Projekten bei fachlicher Definition von Datenqualität und bei bereichsübergreifender Abstimmung ■ Unterstützung des Technischen Designs bei der Implementierung fachlicher Datenqualitätsprüfungen sowie Umsetzung/Freigabe von kurzfristigen Datenkorrekturen ■ Erstellung von Standard- und Sonderauswertungen für Fachbereiche ■ Kontaktperson für Open Government Data (OGD) und Geoinformationssysteme (GIS) 	<ul style="list-style-type: none"> ■ Verantworten Umgang mit den Daten und sind in Dienststellen beziehungsweise Fachbereichen angesiedelt ■ Definieren Anforderungen an Daten sowie bezüglich der Datenqualität und Datenverfügbarkeit im Fachbereich ■ Formulieren den Datenbedarf ihrer Organisationseinheit an Data Steward, prüfen und begleiten die Implementierung daraus abgeleiteter Vorhaben ■ Datenschutzverantwortliche müssen als Data Experts nominiert und qualifiziert sein ■ Fachliche und technische Data User aus den Fachbereichen berichten an Data Expert
Data User	<ul style="list-style-type: none"> ■ Erfassung ■ Auswertung ■ Standardisierung 	<ul style="list-style-type: none"> ■ Einhaltung der Datenqualität ■ Durchführung von Plausibilitätskontrollen ■ Standardauswertungen für eigenen Fachbereich 	<ul style="list-style-type: none"> ■ Produzieren und/oder Nutzen von Daten unter Einhaltung festgelegter Grundsätze ■ Kennen Datenquellen, verstehen verwendete Daten und extrahieren diese für spezifische Bedürfnisse ■ Data User geben Feedback bezüglich der Datenqualität an jeweilige Data Experts

2. Datenstrategien und Datenqualitätsmanagement

Das Datenqualitätsmanagement ist die Basis für die strategische Ausrichtung eines Datenmanagements, denn dieses verbessert die Qualität der Daten, es optimiert somit den Wert der Informationen und dadurch die Nutzbarkeit. Das heißt: Kein Informationsmanagement ohne Datenmanagement. Im Mittelpunkt steht die Konzeption einer langfristigen Strategie, wie Daten zukünftig optimal geplant, erhoben und genutzt werden können – sowohl inhaltlich als auch technisch. Ziel ist die Erhöhung interner Effizienz, Vermeidung von Redundanz und Reduktion von Fehlerquoten.

Mithilfe von Data Governance erfolgt die verbesserte Nutzung der Daten in allen Fachbereichs- und Kernprozessen der kommunalen Verwaltung. Ausgehend von den Zielen des Datenmanagements erfolgt das Operationalisieren auf einer abstrakten Ebene durch die Formulierung einer Datenstrategie. Die Datenstrategie legt die Aktivitätsfelder fest, strukturiert die fachlichen Anforderungen und basiert auf den Vorgaben und Empfehlungen der Data Governance (siehe Kapitel 1). Für die Datennutzung und Abstimmung mit/Abgrenzung von bestehenden Ansätzen der IT-Systemarchitektur (siehe Kapitel 3) hat die Datenstrategie den gesamten Lebenszyklus eines Datums sicherzustellen.

Im Rahmen des Modellprojektes wurden die folgenden thematischen Schwerpunkte und Zielstellungen von der Projektleitung definiert. Diese waren Gegenstand der Leistungsbeschreibung im Vergabeverfahren und als wesentliche Leistungen durch die Auftragnehmer zu erbringen. Das Ziel bestand darin, zu den definierten Fragestellungen konkrete, für die kommunale Praxis der Stadt Cottbus operable Leitfäden zum Umgang mit den städtischen Daten zu entwickeln:

- **Leitfaden Datenplanung** | Es erfolgt eine zeitgerechte und zielgerichtete Planung der Vorhaben, die frühzeitig und systematisch nach Kriterien eines Bedarfsmanagements zu erfassen sind. (Krohn & Heinze GbR)
- **Leitfaden Datenerhebung** | Der Einsatz neuer digitaler Erhebungsmethoden und eine verbesserte Nutzung bestehender administrativer Daten sind anzustreben. Ressourcensparende Methoden zur Produktion von datengesteuerten Informationen sowie Überprüfung von Periodizität und Umfang der Befragungen sind von den Fachbereichen der Stadtverwaltung mit dem zukünftigen „Datenmanagementbereich“ abzustimmen. (Krohn & Heinze GbR)
- **Leitfaden Datenqualität** | Die Einordnung von Qualitätskriterien in systematische Kategorien in Bezug auf Relevanz, Aktualität, Genauigkeit, Vergleichbarkeit und Kohärenz ist darzustellen. (Krohn & Heinze GbR)
- **Leitfaden Datenanalytik** | Das Beherrschen und Analysieren von sehr großen und stetig wachsenden Datenmengen ermöglicht eine formalisierte Informationsgewinnung. Mit Blick auf die Datensicherungsstrategie wird die Umsetzung eines effizienten Datenspeicherungskonzeptes für die Durchführung von Standardauswertungen und vertiefende Untersuchungen nach dem jeweiligen Verwendungszweck forciert. (SyncWork AG)
- **Leitfaden Datenvisualisierung** | Die Anforderungen an die Datenvisualisierungen dienen zur Sicherstellung einer klaren Kommunikation. Datenvisualisierungen dienen dem Präsentieren von Daten, um komplizierte Datensätze mit einer hohen Informationsdichte zu verstehen. (SyncWork AG)
- **Leitfaden Datenschutz und Datensicherheit** | Die kommunale Verwaltung orientiert sich an den geltenden rechtlichen Grundlagen, den Vorgaben von Bund und Land, den internationalen Normen und Standards sowie den internen IT- sowie den allgemeinen Datenschutzrichtlinien. (SyncWork AG)

Angangsbasis sind die von der Krohn & Heinze GbR im Zeitraum von Juli bis September 2020 in den drei Bereichen Team Statistik, Fachbereich Stadtentwicklung sowie Fachbereich Geoinformation und Liegenschaftskataster durchgeführten Ist-Analysen. Hierzu wurden die eingesetzten Fachanwendungen und Programme vor Ort als potenzielle Datenquellen analysiert sowie die bereitgestellten Datenbestände der Fachbereiche hinsichtlich Strukturierung, Qualität, Bedeutung, Metadaten, Menge, Zeitbezug und Aktualisierungsintervall ausgewertet. Mithilfe von Interviews, Fragenkatalogen und Datenexploration wurden die benötigten Angaben ermittelt. Im Ergebnis wurde festgestellt, dass es sich größtenteils um unstrukturierte Daten handelt, die oft nur in Form manueller Exporte als CSV- oder Excel-Dateien bereitgestellt werden können. Einige Datenquellen lassen auch den Zugriff über automatisierbare Schnittstellen wie SQL oder WFS/CSW zu.

Rund 200 Anforderungs- und Zielkriterien, die mit den Projektteilnehmenden aus den drei Fachbereichen erarbeitet wurden, sind im Evaluierungsprozess für das Projektziel Data Warehousing und Datenintegration berücksichtigt worden. Diese wurden nach Zielen, Rahmenbedingungen, Restriktion und Anforderung klassifiziert sowie den einzelnen Aufgaben/Aufgabenbereichen einer Datenintegrationslösung und den Schichten der DWH-Referenzarchitektur zugeordnet. Anschließend wurden diese Kriterien von der Andreas Krohn & Birk Heinze GbR sowie dem CIO der Stadt Cottbus/Chósebus auf einer Skala von 1 bis 10 bewertet, wobei 10 die höchste Gewichtung/Notwendigkeit darstellt.

Tabelle 2: Gewichteter Anforderungskatalog der Fachbereiche der Stadt Cottbus zur städtischen Data Governance
(Quelle: Dr. Gustav Lebhart, Krohn & Heinze GbR)

Anforderung	Gewicht
Excel-Integration (Excel als Datenquelle importieren sowie Excel als Results exportieren)	9,5
Unterstützung DSGVO & Datenschutzbestimmungen/rechtskonformes Hosting nach EU-Datenschutzrecht	9,5
Effiziente Produktion, Mehrfachverwertung und Auffindbarkeit von Informationen	9
Zentrale fachbereichsübergreifende Verwaltung der Berechtigungen und Verantwortlichkeiten (Rollen, gruppenbasiert)	9
Strukturänderungen im Datenmodell dürfen nicht zu Informationsverlust führen	9
Änderungen müssen mit minimalem Aufwand durchführbar und abwärtskompatibel sein	9
Datenquellen sind relationale Datenbanken (Native, ODBC, JDBC) und strukturierte Dateien (Excel, CSV, Shapefiles)	9
Integration überwiegend strukturierter Daten im Core DWH mittels relationalem DBMS	9
Lizenz- und Wartungskosten während des laufenden Betriebes	9
Integriertes Datenmanagement mit einem zentralen „Datenkatalog“ bzw. „Metadaten-Repository“	8,5
Einheitliche Begriffswelt innerhalb der Stadtverwaltung anhand eines gemeinsamen Geschäftsglossars herstellen	8,5
ETL-Tools mit konfigurierbaren, automatisierbaren ETL-Prozessen für Mapping, Datenbereinigung, Datenanreicherung, Validierung, Aggregation	8,5
Unterstützung des zentralen Stammdaten-Repository im DWH	8,5
Einfach zu bedienendes Front-End soll dem Anwendenden die komfortable Durchführung von Abfragen und Auswertungen ermöglichen	8
Historisierte Daten werden zentral im DWH bereitgestellt und nicht in jedem operativen System aufwendig verwaltet. Mit der Einführung des DWH sollen die Daten historisiert werden.	8

2.1. Leitlinie „Datenplanung“

Der Leitfaden skizziert, wie ein Data Warehouse beziehungsweise ein zentrales, harmonisiertes Datenmodell entwickelt werden kann. Hierzu werden unterschiedliche Methodiken sowie deren jeweilige Vor- und Nachteile beschrieben: vom „Top-Down“ gesteuerten Ansatz über ein „Bottom-Up“-Vorgehen bis zu einem hybriden Mittelweg.

Des Weiteren wurde in diesem Leitfaden eine mögliche Reihenfolge für die untersuchten Datenthemen erarbeitet. Sie basiert auf der Erkenntnis, dass circa 50 Prozent der Endnutzenden zur Erledigung ihrer Aufgaben

lediglich den zentralisierten Zugriff auf die „Rohdaten“ benötigen. Erst später folgen die Anforderungen nach konkreten Reports oder bestimmten Kennzahlen der Stadtverwaltung (Key Performance Indikatoren, KPIs).

- Phase1: Fachübergreifend vorkommende Stammdaten zentral bereitstellen
- Phase2: Prozessdaten mit Bezug zu den zentralen Stammdaten für Statistiken, Reports bereitstellen
- Phase3: Individuelle Fachbereichsdaten ergänzen

Ergänzend wurde das Durchführen eines Proof of Concept (POC) vorgeschlagen sowie eine Checkliste hierfür erarbeitet. Der POC soll sowohl die Umsetzbarkeit mit den favorisierten Systemkomponenten als auch die Machbarkeit des vorgeschlagenen Planungsprozesses mit den gewählten Datenthemen verifizieren.

2.1.1. Der Top-Down-Ansatz

Beim anforderungsgetriebenen Top-Down-Ansatz wird ausgehend von den gewünschten Endresultaten, wie benötigte Reports, Kennzahlen, Sichten und die dafür notwendigen Datenzusammenstellungen (Data Marts), ein zentrales Datenmodell entwickelt, welches die Bedarfe optimal erfüllen soll. Das sind im Wesentlichen:

- Anforderungen der Fachanwendenden wie gewünschte Analysen, Reports und Sichten auf die Daten
- Anforderungen aus der Data Governance-Richtlinie der Kommune als zentrale Vorgaben

Ergebnis dieser Vorgehensweise ist oft ein am dimensionalen Datenmodell der gewünschten Data Marts ausgerichtetes Datenmodell mit Stern- oder Schneeflockenschema.

Der Top-Down-Ansatz ist dann zu empfehlen, wenn die Anforderungen klar definiert sind und nur wenige beziehungsweise mehrere ähnliche Datenzusammenstellungen (Data Marts) für die gesuchten Ergebnisse benötigt werden. Bei Vorhaben mit stark unterschiedlichen fachlichen Abteilungen, sehr ungleichen Quelldaten oder nur vage bekannten Anforderungen ist der Ansatz aufgrund des hohen Aufwandes bei nachträglichen Änderungen nicht zu empfehlen.

Tabelle 3: Vor- und Nachteile des Top-Down-Ansatzes bei der Erstellung des zentralen Datenmodells für das Core Data Warehouse (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Das Datenmodell ist vollständig bezüglich der initialen Anforderungen.	Sehr hoher initialer Aufwand, da alle Anforderungen im Vorfeld umfassend ermittelt, analysiert und für alle Beteiligten verständlich formuliert werden müssen.
Die Data Governance-Anforderungen sind frühestmöglich berücksichtigt worden.	Unflexible Datenmodelle führen zu hohem Aufwand bei nachträglichen Änderungen, und zwar sowohl bei den Änderungen am Datenmodell als auch bei Änderungen der Transformationsprozesse.
Die Daten im Core Data Warehouse (Core DWH) entsprechen den aktuell definierten Geschäftsregeln (Schema on Write).	Die Transformation der Daten in das Core Data Warehouse (Core DWH) kann sehr komplex werden, wenn Struktur und Qualität der Quellsysteme stark vom zentralen Datenmodell abweichen.
Das Datenmodell des Core DWH ist an den benötigten Datenzusammenstellungen der Ergebnisse (Data Marts) orientiert und erlaubt in der Regel einfache Datentransformationen vom Core DWH in die Data Marts.	Es kann vorkommen, dass die geforderten Daten durch die Quellsysteme gar nicht geliefert werden können und andere Datenquellen hinzugezogen werden müssen.
Der Core DWH enthält nur die gemäß den Anforderungen benötigten Daten und ist daher möglichst schlank und übersichtlich gehalten.	Anforderungen sind ständigen Änderungen unterworfen, insbesondere in Bezug auf neue Datenquellen und Benutzerkreise oder geänderte Interessen und Ziele.

Vorteile	Nachteile
Der Ansatz ist ausgerichtet auf eine stark zentralisierte Vorgehensweise von oben durch ein kleines Expertenteam.	Änderungen am zentralen Modell ziehen umfangreiche Tests und Abwandlungen der bisherigen Transformationsprozesse nach sich. Daten, die bisher nicht relevant waren, müssen bei Bedarf nachträglich geladen und über Quellsystemarchive historisiert werden.
	Eine nachträgliche Historisierung ist nur mit sehr großem Aufwand aus den Datenarchiven der Quellsysteme durchführbar.

2.1.2. Der Bottom-Up-Ansatz

Beim datengetriebenen Bottom-Up-Ansatz wird ausgehend von den zu integrierenden Quellsystemen ein zentrales Datenmodell entwickelt, welches die unterschiedlichen Daten der Quellsysteme historisiert, verknüpft und harmonisiert in einheitlichen Strukturen ablegen soll. Die Hauptmotivation umfasst im Wesentlichen:

- Das generelle Bereitstellen aller Daten der Quellsysteme an einem zentralen Ort
- Das Zusammenführen und Harmonisieren der Daten
- Das Historisieren der Daten zum Zweck zeitlicher Auswertungen

Ergebnis dieser Vorgehensweise ist oft ein streng normalisiertes, sehr feingranulares Datenmodell, welches an den stark normalisierten Datenmodellen der operativen Quellsysteme ausgerichtet ist.

Der Bottom-Up-Ansatz ist dann zu empfehlen, wenn die Anforderungen noch unvollständig definiert sind und der Fokus auf der Sammlung und schnellen zentralen Bereitstellung aller verfügbaren Daten in historisierter Form liegt. Hiermit wird oft schon ein Großteil der Anwendenden bezüglich ihrer Anforderungen gut bedient.

Tabelle 4: Vor- und Nachteile des Bottom-Up-Ansatzes bei der Erstellung des zentralen Datenmodells für das Core Data Warehouse (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Das Datenmodell ist an den Quellsystemen orientiert und ermöglicht daher einfachere Ladeprozesse.	Für ein gut zu integrierendes und zu harmonisierendes zentrales Datenmodell müssen alle Quellsysteme im Vorfeld analysiert werden, was einen höheren initialen Aufwand bedeutet und eine hohe Komplexität erreichen kann.
Die verfügbaren Daten können frühzeitig in das Core DWH geladen und historisiert werden.	Ohne diesen initialen Aufwand erreicht man häufig nur isolierte 1:1 Abbildungen der Quellsysteme mit zusätzlicher Historisierung. Die echte Integration, Harmonisierung und Erfüllung der Geschäftsregeln findet dann erst bei der Erstellung und Befüllung der gewünschten Data Marts durch aufwendige Transformationen statt.
Die Quelldaten werden in ihrer feinsten Granularität in das Core DWH geladen.	Ohne umfassend analysierte Anforderungen ist es schwer, die richtige Datenstruktur für das zentrale Datenmodell zu entwickeln.
Die vollständigen und historisierten Daten der einzelnen Fachbereiche ermöglichen eine schnelle Darstellung einfacher Sichten und Reports aus den bereitgestellten Daten.	Data Governance-Anforderungen gelangen erst spät in den Prozess und damit nur bedingt in das zentrale Datenmodell. Dies führt entweder zu qualitativ minderwertigen Daten oder aufwendigen Integrations- und Transformationsprozessen bei der Erstellung der tatsächlich geforderten Data Marts.
	Sind die Data Governance-Anforderungen nicht im zentralen Datenmodell enthalten, müssen sie bei jeder Erstellung eines Data Marts berücksichtigt werden, was zu komplexen Datentransformationen führt.

Vorteile	Nachteile
	Änderungen am zentralen Datenmodell sind auch bei diesem Ansatz sehr aufwendig, da die existierenden Daten und Transformationen bezüglich der Änderungen geprüft und ggf. angepasst werden müssen.

2.1.3. Der Hybrid-Ansatz am Beispiel von Data Vault

Bei einem hybriden Ansatz wird versucht, die Nachteile der zuvor genannten gegensätzlichen Ansätze durch die Einführung weiterer Datenhaltungsschichten im Core DWH zu eliminieren. Die einzelnen Schichten besitzen klar abgegrenzte Aufgaben. Im Fall des Data Vault-Ansatzes wird die Core DWH-Schicht, wie auf der Abbildung 1 dargestellt, in den Raw Vault und den Business Vault aufgeteilt. Der Raw Vault dient der Aufnahme und Bereitstellung aller Rohdaten in historisierter Form und kann sehr einfach gemäß dem Bottom-Up-Vorgehen erstellt werden. Im Business Vault werden die Daten in dem zentralen Datenmodell anhand der aktuellen Geschäftsregeln harmonisiert und zusammengeführt, wobei sowohl die Bottom-Up- als auch die Top-Down-Vorgehensweise möglich ist. Diese Aufteilung ermöglicht einfachere, modellbasierte und damit automatisierbare Transformationsprozesse von Schicht zu Schicht.

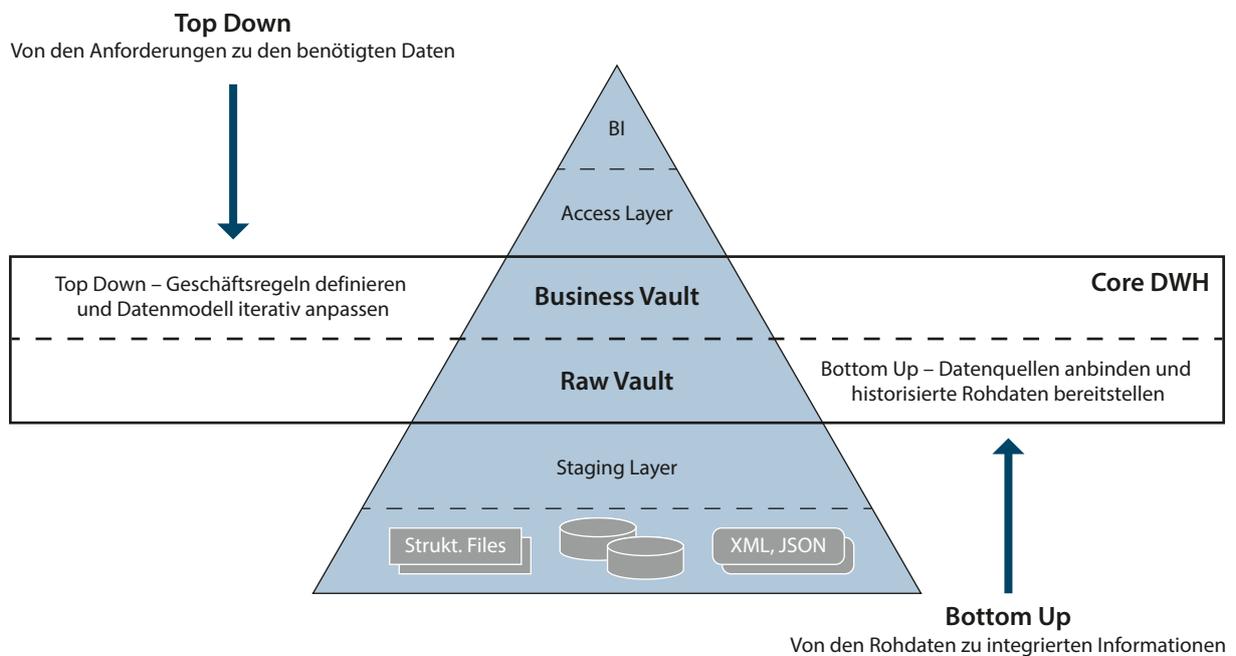


Abbildung 1: DWH-Schichten nach Data Vault (Quelle: Krohn & Heinze GbR)

Der hybride Ansatz bietet die folgenden Vorteile:

- Der hybride Ansatz ermöglicht die einfachere und schnellere Vorhaltung und Bereitstellung (aller relevanten) Rohdaten in historisierter Form in der Bottom-Up-Vorgehensweise.
- Gleichzeitig kann in der Top-Down-Vorgehensweise von den Anforderungen her und mit Blick auf die vorhandenen historisierten Daten das zentrale Datenmodell mit den aktuellen Geschäftsregeln entwickelt werden. Ändern sich die Geschäftsregeln, kann das zentrale Datenmodell aus den historisierten Daten neu befüllt werden.

- Der hybride Ansatz erlaubt die Verteilung der Aufgaben auf mehrere Benutzerrollen im Integrationsteam. Er ist nicht mehr ausschließlich IT-zentriert und bietet Möglichkeiten zur parallelen Abarbeitung durch mehrere Gruppen.

2.2. Leitlinie „Datenerhebung“

Auch die Leitlinie „Datenerhebung“ wurde auf der Basis einer Auswertung der aktuellen Datenflüsse und Bestände der Stadt Cottbus/Chósebus im Dialog mit den beteiligten Fachbereichen (s. o.) entwickelt. Im Zuge der Analyse des Ist-Zustandes wurde deutlich, dass die Datenflüsse der Cottbuser Stadtverwaltung nur schwer administrierbar sind, weil Datenquellen oft dateibasiert und damit stark fragmentiert zur Verfügung stehen. Fachübergreifende Kooperationen finden statt, jedoch erfolgen diese teilweise mit ungeeigneten Methoden wie manuelle Anfrage und Dateiaustausch per E-Mail oder über ein freigegebenes Verzeichnis im Fachbereich.

Viele Daten liegen nur lokal und knapp dokumentiert in den Fachbereichen vor und sind für andere Fachbereiche nur schwer zugänglich.

Das Geoportal der Stadt Cottbus/Chósebus stellt eine Auswahl öffentlich zugänglicher Datenthemen geodatenbasiert über die digitale Stadtkarte zur Verfügung und bietet eine INSPIRE-konforme Bereitstellung dieser Daten an. Die ausgewählten fachlichen Datenthemen liegen hier zentral gesammelt und georeferenziert vor. Der Fokus liegt auf der Vereinheitlichung der Geodaten, inklusive der hierzu notwendigen Metadaten.

Es existiert kein fachbereichsübergreifendes, einheitliches Management- und Freigabekonzept für Daten und Dateien. Zudem ist noch kein zentraler und einheitlicher Übergabemechanismus implementiert. Dies ist jedoch eine Voraussetzung, um effizient den zu administrierenden Transfer von Daten aus einer Quelle zu einem Ziel zu gestalten und uneinheitliche, individuelle und damit sehr zeit- und kostenaufwendige Managementprozesse zu vermeiden.

2.2.1. Zielarchitektur

Ausgehend von dem Ist-Zustand und den daraus resultierenden Data Governance-Herausforderungen sowie den initial formulierten Anforderungen konnten folgende Hauptziele formuliert werden.

Tabelle 5: Hauptziele der Datenerhebung (Quelle: Krohn & Heinze GbR)

Anforderungen an die Datenerhebung	
Zentrales Quellenverzeichnis	Eine zentral organisierte und damit administrierbare Verwaltung der Datenquellen über ein zentrales Quellenverzeichnis. Hierzu gehören eine zentrale Katalogfunktion zur Auflistung und Beschreibung der Datenquellen anhand von Metadaten sowie eine zentral organisierte Ablage, die Landing Zone, der für die Datenintegration gelieferten Quelldateien.
Zentrales Data Warehouse	Datenintegration, also das Sammeln, Historisieren, Zusammenführen und Konsolidieren der heterogenen Daten einzelner Fachabteilungen über ein zentrales Datenmodell führt zum gesamtheitlichen Blick auf die unternehmensweiten Daten. Im zentralen Data Warehouse werden diese integrierten Daten dauerhaft gespeichert sowie zur Analyse und Auswertung bereitgestellt.
Eigenständige Fachanwendungen	Die Fachanwendungen bleiben von der Einführung der Datenintegrationsschicht unabhängig. Die operativen Systeme sind durch das zentrale DWH von der Analyse- und Reporting-Software entkoppelt, Datenauswertungsprozesse belasten diese somit nicht mehr. Über Schnittstellen kann jedoch auf bereinigte und harmonisierte unternehmensweite Stammdaten zugegriffen werden, was deren Qualität in den Fachanwendungen verbessern kann.

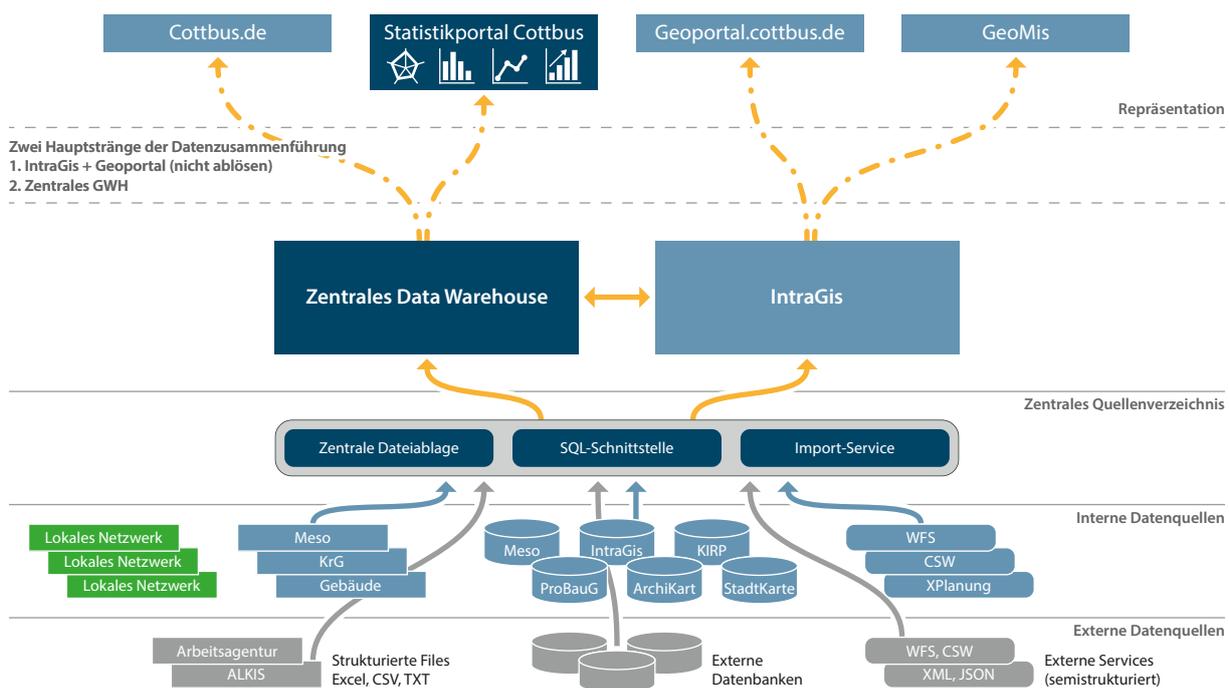


Abbildung 2: Zentrales Data Warehouse als Integrationsschicht (Quelle: Krohn & Heinze GbR, 2020)

Abbildung 2 zeigt schematisch die Zuordnung der evaluierten vorhandenen internen Fachanwendungen und Datenquellen (hellblau dargestellt) sowie der neu zu schaffenden Systemkomponenten (dunkelblau dargestellt) zu den einzelnen Schichten eines Datenintegrationssystems. Dies schließt die Anbindung externer Datenquellen und Anwendungen (grau dargestellt) ein. Ein zentrales Quellenverzeichnis verwaltet die Metadaten für die anzubindenden Datenquellen inklusive der verfügbaren Schnittstellen. Das zentrale Data Warehouse bildet die Integrationsschicht und stellt die Daten harmonisiert und aufbereitet für das Statistikportal und weitere Abnehmer bereit. Das vorhandene Geoinformationssystem (IntraGis) kann sowohl als Quelle für das DWH dienen als auch neue Datenthemen aus dem DWH beziehen.

2.2.2. Anforderungen an das Datenmanagement

Das Ergebnis einer gewachsenen Landschaft von Applikations- und Datensilos ist eine Datenfragmentierung. Jede Applikation hat eigene Meta- und Stammdaten. Diese Tatsache erschwert die Durchführung von Verwaltungstätigkeiten und kann weitere Kosten und Ressourcen verursachen. Die nachfolgende Tabelle charakterisiert die Herausforderungen des Datenmanagements.

Tabelle 6: Herausforderungen des Datenmanagements (Quelle: Krohn & Heinze GbR)

Anforderungen an das Datenmanagement	
Vermeidung redundanter Daten	Datenredundanz ist ein Kostentreiber und erfordert eine aufwendige Pflege der Daten über die Applikationsinseln hinweg. Bei jeder Datenänderung oder -neueingabe ist sicherzustellen, dass die Änderungen in allen betroffenen Datenbanken erfolgen.
Vermeidung inkonsistenter Datenstrukturen	Applikationen administrieren ihr eigenes Datenvokabular gemäß der spezifischen Aufgabenstellung. Datenmerkmale und deren Ausprägungen unterscheiden sich von einer Fachapplikation zur anderen.

Anforderungen an das Datenmanagement	
Vermeidung widersprüchlicher Daten	Datenredundanz und Dateninkonsistenz sorgen für ein unterschiedliches Verständnis von Daten. Die verschiedenen Organisationseinheiten der öffentlichen Verwaltung definieren und bewerten die Daten in ihren Applikationen unterschiedlich. So entsteht ein gravierendes Datenqualitätsproblem.
Verfügbarkeit der Daten durch Integration	Daten, die für eine Auswertung benötigt werden, können aus den vorhandenen Fachapplikationen nicht entnommen werden. Das verursacht in der Regel Zusatzkosten, da ein spezielles Bindeglied oder eine Schnittstelle für den Datenzugriff erforderlich ist und ein Budget für die Lizenzierung und Pflege dieser Komponenten bereitzuhalten ist.

2.3. Leitlinie „Datenqualität“

Datenqualität bezeichnet die Bewertung von Datenbeständen in Bezug auf die Eignung zur Erfüllung spezifischer Anforderungen oder/und Standards von Anwendenden – kurz: „Wie geeignet sind die Daten, um einen spezifischen Zweck zu erfüllen?“ Dabei geht es insbesondere um die Korrektheit, Relevanz, Verlässlichkeit, Konsistenz und zuletzt um die Verfügbarkeit der Informationen.

Tabelle 7: Anforderungen an das Datenqualitätsmanagement (Quelle: Krohn & Heinze GbR)

Anforderungen an das Datenqualitätsmanagement	
Datenqualität	Das Erreichung oder die Gewährleistung einer angestrebten Datenqualität ist ein iterativer, dynamischer und kontinuierlicher Verbesserungsprozess, der im Rahmen einer Data Governance-Strategie in der Kommune verankert sein muss. Die Einbindung geeigneter Datenqualitätskontrollen in die Datenverarbeitungskette ist entscheidend für den Erfolg der Data Governance-Initiative.
Datenintegrität	„Datenintegrität beschreibt die Richtigkeit, Vollständigkeit und Konsistenz der Daten über die gesamte Lebensdauer. Das bedeutet auch, dass Daten nicht unerkannt verändert werden dürfen und der Fokus auf der Nachvollziehbarkeit von Datenänderungen liegt“ (Spilka, 2019) ¹ . Hierzu gehören neben dem Aspekt der Datenqualität auch die Datensicherheit sowie die Einhaltung von Datenschutzvorgaben wie beispielsweise die Vorgaben der Datenschutz-Grundverordnung (DSGVO).
Qualitätskriterien	Zur Bewertung der Datenqualität können verschiedene Kriterien herangezogen werden. Die Gewichtung der einzelnen Kriterien kann je nach Anwendungsfall, Aufgabe und Datenschicht variieren. Sind zum Beispiel Eindeutigkeit, Einheitlichkeit, Redundanzfreiheit und Konsistenz der Daten innerhalb eines einzelnen operativen Systems gewährleistet, kann es bei der Datenintegration mehrerer operativer Systeme zu Verletzungen dieser Kriterien kommen.
Qualitätsmessung	Durch Geschäftsregeln werden die Vorgaben, Bedingungen und Anforderungen formuliert, welche die logische Konsistenz der Daten definieren und als Metadaten erfasst werden. Als Kennzahl kann der „Erfüllungsgrad“ je Regel „in Prozent“ definiert werden, wobei die untersuchte Menge sich entweder auf eine Anzahl an Datensätzen, Attributen, Tabellen oder individuellen Datenobjekten beziehen kann. Insbesondere bei Bestandsdaten können Regelabweichungen entstehen, wenn die Geschäftsregeln nachträglich formuliert werden oder die Validierung der Regeln nicht beziehungsweise nur unvollständig erfolgt.

Das Thema Datenqualität zentral im Unternehmen zu etablieren ist eine der Hauptaufgaben einer Data Governance-Strategie. Hierzu gilt es, Datenqualitätsrichtlinien zu definieren, die Mitarbeitenden zu befähigen und die notwendige Software bereitzustellen.

¹ Spilka, S. DATENINTEGRITÄT: BEDEUTUNG, RISIKEN & DER RICHTIGE UMGANG.
Internetquelle: <https://anexia.com/blog/de/datenintegritaet/> (3. Oktober 2019).

Ziel des Datenqualitätsmanagements ist, Datenformate zu vereinheitlichen, Redundanzen aufzulösen und in eine gemeinsame Struktur zu bringen. Auch eine allgemeingültige Definition von Kennzahlen und deren Merkmalsausprägungen, zum Beispiel welche Bevölkerungsgruppen und räumlichen Bereiche unterschieden werden, in welchen Zeitabständen eine Analyse erfolgt oder welche Kennzahlen miteinander kombiniert betrachtet werden, ist in Bezug auf die Definition von Bedeutung.

Die Übergabepunkte von Daten-Layer zu Daten-Layer in einer mehrschichtigen DWH-Architektur stellen die geeigneten Qualitätsmesspunkte dar. Hier werden die Daten jeweils aus einer Quelle extrahiert, validiert und bereinigt sowie abschließend in die Zielstruktur transformiert. In den einzelnen Schichten können unterschiedliche Aspekte der Datenqualität umgesetzt werden (siehe Abbildung 3).



Abbildung 3: Aspekte der Datenqualität je Datenhaltungsebene (Quelle: Krohn & Heinze GbR)

Während es bei den Datenquellen darum geht, zunächst die relevanten Daten zu ermitteln und auszuwählen, können nach der Extraktion in die Landing Zone weitere Analysen in Bezug auf die Qualität der extrahierten Daten erfolgen. In jeder folgenden Schicht werden zusätzliche Geschäftsregeln ergänzt, um die Daten über ein gemeinsames Datenmodell zu integrieren und zu vereinheitlichen. Letztlich werden im Access Layer sehr fachspezifische Regeln und Strukturen definiert, um die Fachanforderungen der Anwender zu erfüllen.

Bei jedem Transformationsschritt von Schicht zu Schicht muss nachvollziehbar festgehalten werden, welche Daten erfolgreich transformiert und welche Daten als fehlerbehaftet markiert beziehungsweise gänzlich zurückgestellt wurden.

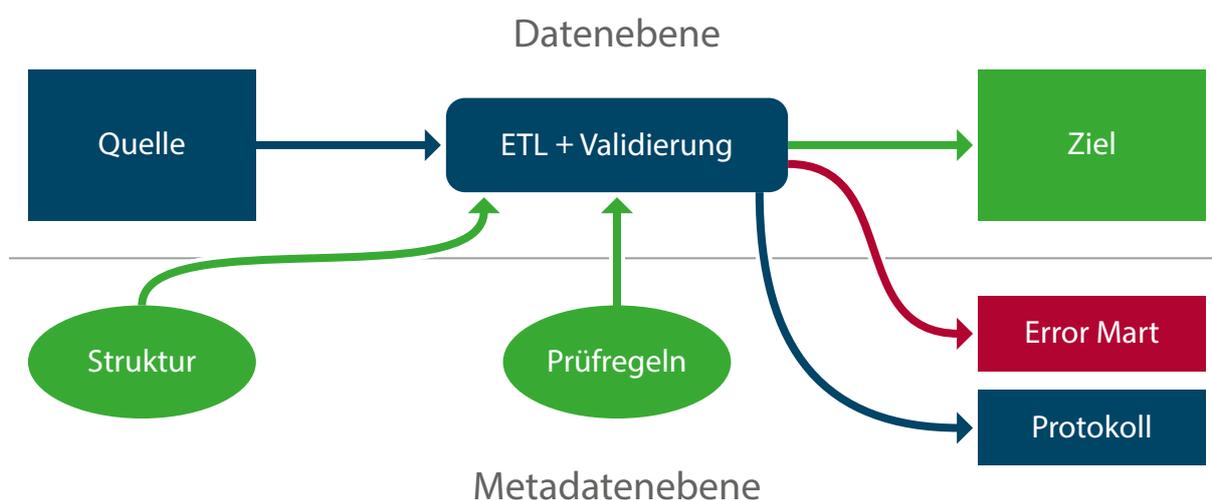


Abbildung 4: Operative Metadaten nach Data Vault (Quelle: Krohn & Heinze GbR)

Beim Data Vault-Ansatz ist hierfür das Konzept des Error Mart vorgesehen (siehe Abbildung 4). Error Marts sind spezifische Information Marts zur Aufnahme und Bereitstellung der Fehlerreports und der betroffenen Datensätze. Sie werden als operative Metadaten bereitgestellt.

2.4. Leitlinie „Datenanalytik“

Für die Erstellung dieses Leitfadens wurden die Teilnehmenden aus den Fachbereichen Stadtentwicklung, Geoinformation und Liegenschaftskataster sowie Statistik der Stadtverwaltung Cottbus/Chósebusz mittels Online-Workshops, halbstrukturierter Interviews und qualitativer Fragebögen zu der Ausgangssituation sowie den Anforderungen an ein Informationsportal zur Datenanalytik befragt. Ihre Einschätzungen und Antworten sind in die Darstellungen dieses Leitfadens eingeflossen.

Datenauswertungen haben das Ziel, aus den vorhandenen Rohdaten unter Anwendung verschiedener Methoden und statistischer Analyseverfahren Informationen und Erkenntnisse zu gewinnen, zu beschreiben und darzustellen. Im Rahmen der statistischen Auswertung werden die Daten geordnet und strukturiert. So können beispielsweise Muster oder Trends in den Daten gefunden werden. Die Methodik und die angewandten Verfahren werden von der Fragestellung sowie der Art und Qualität der Daten bestimmt.

Analysen von kommunalen Daten sind entscheidend für die regelmäßige Beobachtung der kommunalen Entwicklung in den Feldern Demografie, Wirtschaftsentwicklung, Bildung, Soziales, Verkehr, Gesundheit und weiterer kommunaler Wirkungsbereiche. Nur eine valide Datenanalyse kann dabei Grundlage einer entscheidungsbasierten Kommunalpolitik sowie eines effizienten und zielgenauen Einsatzes kommunaler Ressourcen sein.

Tabelle 8: Kriterien der Datenauswertung (Quelle: SyncWork AG)

Kriterien bei der Datenauswertung	
Fragestellung und Planung	In der ersten Phase des Analyseprozesses geht es darum, die Fragen zu definieren, die mithilfe der ausgewerteten Daten zu beantworten sind und in einem Ideenfindungsprozess mit anderen Fachbereichen und dem Kollegium präzisiert werden können. Je individueller die Fragen sind, desto spezifischer können die Anforderungen erläutert werden.
Datenerhebung und Datenaufbereitung	Zunächst sind die gesuchten Daten und Fakten zu identifizieren. Anschließend sind die Datenquellen mit den Rohdaten für diese gesuchten Daten festzulegen. Es ist sicherzustellen, dass aus diesen Rohdaten die relevanten absoluten Kennzahlen oder Indikatoren für die Datenanalysefragen abgeleitet werden können. Bei der Datenaufbereitung sind die Daten in das richtige Format für die nachfolgende Datenanalyse zu bringen und gegebenenfalls nutzlose Daten auszuklammern.
Durchführung der Datenanalyse	Bei der Datenanalyse kommen statistische Methoden zur Anwendung, um aus Quelldaten relevante Information zu gewinnen, aus denen sich Entscheidungsgrundlagen ableiten lassen. Daraus ergeben sich Anforderungen, die die Daten erfüllen müssen. Sie sollen: <ul style="list-style-type: none"> ■ leicht zu analysieren sein, ■ ansprechend visualisiert werden können und vor allem ■ zuverlässig und valide sein.
Datenpräsentation	Bei der Datenpräsentation sollte die korrekte Informationsvermittlung im Vordergrund stehen. Die Übersichtlichkeit und Klarheit haben oberste Priorität. Besonderes Augenmerk muss auf die Wahl des passenden Diagrammtyps gelegt werden. Nicht jeder Diagrammtyp ist für alle Darstellungen geeignet. Bei der Erstellung und Formatierung von Tabellen ist es wichtig, die präsentierten Daten klar und eindeutig strukturiert sowie optisch ansprechend aufzubereiten.
Dateninterpretation	Die Dateninterpretation weist den Ergebnissen aus der Datenanalyse eine Bedeutung zu und hat zum Ziel, dass die kommunale Verwaltung die richtigen Schlussfolgerungen beziehungsweise die politischen Verantwortlichen die entsprechenden Konsequenzen ziehen können. Letztendlich ist die Dateninterpretation der entscheidende letzte Schritt für den Erfolg der Analyse und muss sorgfältig durchgeführt werden.

Wichtigstes Ziel der Datenanalyse ist, aus den vorhandenen Daten durch Anwendung verschiedener Methoden und Analyseverfahren die jeweils benötigten Erkenntnisse oder Informationen zu gewinnen. Da es kein Standardvorgehen bei der Durchführung von Datenanalytik im kommunalen Bereich gibt, muss jede Kommune ihre eigenen Standards, Regeln und Prozesse im Bereich der Data Governance festlegen. Die technische Umsetzung der Data Governance findet mittels des Datenmanagements statt, welches alle Maßnahmen, die zur Erhebung, Bereitstellung, Verwendung und Qualitätssicherung von Daten notwendig sind, erfasst. Zu den Maßnahmen gehören vor allem die regelhafte Zusammenführung, Standardisierung und Pflege erhobener Daten. Die Datenqualität wird durch die Merkmale Gültigkeit, Genauigkeit und Vollständigkeit bestimmt.

Eine Data Governance definiert im Zusammenhang mit der Datenerhebung und -verarbeitung Strukturen mit Verantwortlichen, Richtlinien, Regeln, Prozesse und Kriterien für den gesamten Lebenszyklus von Daten. Wichtig ist in diesem Zusammenhang die Festlegung einer eindeutigen Verantwortlichkeit für die Pflege der Daten.

2.5. Leitlinie „Datenvisualisierung“

Datenanalysen sind in der Regel abstrakt und nicht selten liegen Ergebnisse in Form von relativ unüberschaubaren Excel-Tabellen vor. Im Vergleich dazu haben Datenvisualisierungen Vorteile: Die zentralen Aussagen und Zusammenhänge lassen sich meist in wenigen Augenblicken erfassen und die Darstellung erweckt Vertrauen, zumal diese Datenvisualisierungen auf verifizierten Zahlen und Fakten basieren sollten. Datenvisualisierungen sind transparentere und verständlichere Kommunikationsformen.

Datenvisualisierungen haben einen besonderen Stellenwert bei der Vermittlung von evidenzbasierten Daten. Bilder und Grafiken stellen einen wichtigen Bestandteil der Informationsvermittlung dar. Die Herausforderungen bei der Veranschaulichung von Daten sind daher vielfältig. Da Datenvisualisierungen stets für einen bestimmten Zweck erstellt werden – in der Regel, um von einem bestimmten Sachverhalt zu überzeugen – hat die öffentliche Verwaltung darauf zu achten, dass es zu keinen bewussten oder unbewussten Manipulationen kommt. Mit Datenabbildungen sollen eindeutige Sachverhalte vermittelt werden, die nach bestimmten Darstellungsregeln erstellt werden.

Die Datenvisualisierung bereitet Daten und statistische Auswertungen auf und stellt sie grafisch dar. Wichtige Informationen können schneller erkannt und besser verstanden werden. Mit einem Datenvisualisierungstool stehen unterschiedliche Darstellungsoptionen zur Verfügung. Je nach Datenform kann somit eine geeignete Möglichkeit gewählt werden, um die gewünschten Informationen sinnvoll und nachvollziehbar darzustellen. Prinzipiell werden Daten mit Diagrammen visualisiert. Um komplexe Zusammenhänge vieler Daten übersichtlich und selbsterklärend abzubilden, werden verschiedene Schaugrafiken in (meist interaktiven) Bedienfeldern, sogenannten Dashboards, zusammengestellt. Dieser Leitfaden dokumentiert im Folgenden die wichtigsten Datenvisualisierungsmöglichkeiten mittels grundlegender Diagrammart und Möglichkeiten, die moderne Dashboards bieten.

2.5.1. Gestaltungsebenen der Datenvisualisierung

Für die Datenvisualisierung wird häufig zwischen zwei Gestaltungsebenen unterschieden:

- der grundlegenden Layout-Ebene (Diagrammart) und
- der Zeichen-Ebene (zum Beispiel Farbe, Form), die in der Fachliteratur häufig als retinale beziehungsweise visuelle Variable beschrieben wird.

Um zu einer intuitiv erschließbaren Wiedergabe zu gelangen, kann anhand der Auswahlmatrix für Diagrammart ein geeignetes Layout gewählt oder entsprechend der Auswahlmatrix für retinale Variablen eine geeig-

nete Darstellungsform auf der Zeichenebene gesucht werden, um auch damit eine geeignete Diagrammart zu wählen (welche die jeweiligen Zeichen bietet) beziehungsweise die Visualisierungsform im schon gewählten Schaubild zu optimieren.

2.5.1.1. Auswahl-Matrix: Diagrammartentypen

Wenn bekannt ist, welche Daten visualisiert werden, bietet diese Matrix eine Übersicht von Diagrammartentypen, hierarchisch sortiert nach ihrer Eignung. Allerdings muss hier einschränkend festgehalten werden, dass die Eignung einer Diagrammart immer abhängig von ihrem Inhalt und dem Verwendungszusammenhang ist, der nie allgemeingültig vorweggenommen werden kann.

Tabelle 9: Überblick verschiedener Diagrammartentypen (Quelle: SyncWork AG)

Datenart	Dimension 1-dimensional (1D)	2-dimensional (2D)	multidimensional (XD)
Graphdaten	<ul style="list-style-type: none"> Entscheidungsbaum 	<ul style="list-style-type: none"> Netzwerkdigramm Entscheidungsbaum Treemap 	<ul style="list-style-type: none"> Netzwerkdigramm Entscheidungsbaum
Rangfolgen	<ul style="list-style-type: none"> Tabellen 	<ul style="list-style-type: none"> Balkendiagramm Trichterdiagramm Treemap Wasserfalldiagramm Butterfly-Diagramm Gantt-Diagramm Histogramm Table Heatmap Tabellen Schlagwortwolke 	<ul style="list-style-type: none"> Balkendiagramm Butterfly-Diagramm Table Heatmap Tabellen Schlagwortwolke
Räumliche Daten		<ul style="list-style-type: none"> Karten Heatmap Netzwerkdigramm 	<ul style="list-style-type: none"> 3D-Oberflächendiagramm Netzwerkdigramm
Relationale Daten (nicht-zeitliche Korrelationen, Zusammenhänge mehrerer Variablen, Größenvergleiche)	<ul style="list-style-type: none"> Skala Tachometer-Diagramm Kreisdiagramm 	<ul style="list-style-type: none"> Säulendiagramm Kurvendiagramm Stapeldiagramm Linien-Flächendiagramm Punktvolke Balkendiagramm Trichterdiagramm Sterndiagramm Portfoliodiagramm Butterfly-Diagramm Wasserfalldiagramm Table Heatmap Histogramm Box-Whisker-Plot Tabellen Schlagwortwolke 	<ul style="list-style-type: none"> Portfoliodiagramm Heatmap 3D-Säulendiagramm Säulendiagramm Kurvendiagramm Stapeldiagramm Balkendiagramm Punktvolke Linien-Flächendiagramm Sterndiagramm Butterfly-Diagramm Table Heatmap Tabellen Schlagwortwolke
Temporale Daten (zeitliche Verläufe)		<ul style="list-style-type: none"> Kurvendiagramm Linien-Flächendiagramm Punktvolke Säulendiagramm Stapeldiagramm Portfoliodiagramm Butterfly-Diagramm Gantt-Diagramm Box-Whisker-Plot Table Heatmap Tabellen 	<ul style="list-style-type: none"> Kurvendiagramm Punktvolke Portfoliodiagramm Heatmap 3D-Säulendiagramm Stapeldiagramm Säulendiagramm Linien-Flächendiagramm Butterfly-Diagramm Table Heatmap Tabellen
Textuelle Daten (Text-, Wortdarstellungen)	<ul style="list-style-type: none"> Schlagwortwolke Tabellen 	<ul style="list-style-type: none"> Netzwerkdigramm Schlagwortwolke Tabellen 	<ul style="list-style-type: none"> Netzwerkdigramm Schlagwortwolke Tabellen

Datenart	Dimension		
	1-dimensional (1D)	2-dimensional (2D)	multidimensional (XD)
Zusammensetzung (Aufteilung)	<ul style="list-style-type: none"> ■ Skala ■ Tachometer-Diagramm 	<ul style="list-style-type: none"> ■ Ringdiagramm ■ Sunburst-Diagramm ■ Treemap ■ Stapeldiagramm ■ Wasserfalldiagramm ■ Sterndiagramm ■ Histogramm 	<ul style="list-style-type: none"> ■ Stapeldiagramm ■ Sunburst-Diagramm ■ Linien-Flächendiagramm ■ Sterndiagramm

2.5.1.2. Auswahl-Matrix: visuell-retinale Variablen

Zu visuell-retinalen Variablen gibt es bereits zahlreiche Forschungsarbeiten, die im Wesentlichen auf den Grundlagen bekannter Semiologen (wie zum Beispiel Jacques Bertin beziehungsweise dem Tableau-Designer Jock Mackinlay) basieren. Die Autorenschaft dieses Leitfadens bezieht sich mit der hier dargestellten Matrix auf Empfehlungen von Jock Mackinlay.

Wenn bekannt ist, welche Daten veranschaulicht werden sollen, bietet diese Matrix eine Übersicht über die visuell-retinalen Variablen nach ihrer Eignung sortiert. Je nach Datentyp finden sich die am besten geeignete Variablen ganz oben und sind nach unten hin zwar anwendbar, aber weniger geeignet für die Visualisierung.

Tabelle 10: Übersicht über die visuell-retinalen Variablen, nach ihrer Eignung absteigend sortiert (Quelle: SyncWork AG)

Eignung visuell-retinale Variablen (Farbdichte, Farbe, Fläche, Form, Länge, Neigung, Position, Sättigung, Winkel)		
Quantitativ	Ordinal	Nominal
Position	Position	Position
Länge	Farbdichte	Farbe
Winkel	Sättigung	Farbdichte
Neigung	Farbe	Sättigung
Fläche	Länge	Form
Farbdichte	Winkel	Länge
Sättigung	Neigung	Winkel
Farbe	Fläche	Neigung
Form	Form	Fläche

Quelle: Syncwork AG, 2021, eigene Darstellung in Anlehnung an Mackinlay, 1986, S. 125 ²

2.5.2. Diagrammarten

Die folgenden Ausführungen bieten eine Übersicht über die Diagrammarten. Dabei ist zu berücksichtigen, dass einzelne Diagrammtypen durchaus miteinander kombiniert werden können. Eine Kombination ist vor allem dann zielführend, wenn multidimensionale Daten abgebildet werden sollen.

Diese Übersicht stellt die wichtigsten und gebräuchlichsten Diagrammarten und weitere Visualisierungshinweise dar, kann aber nicht erschöpfend alle bekannten Diagrammarten, die insbesondere für sehr spezifische Anwendungszwecke entwickelt wurden, dokumentieren. Moderne Software-Tools bieten die Möglichkeit, nach Auswahl der jeweiligen Diagrammart mit wenigen Klicks zahlreiche Diagramm-Unterarten für eine geeignete Illustration auszuprobieren.

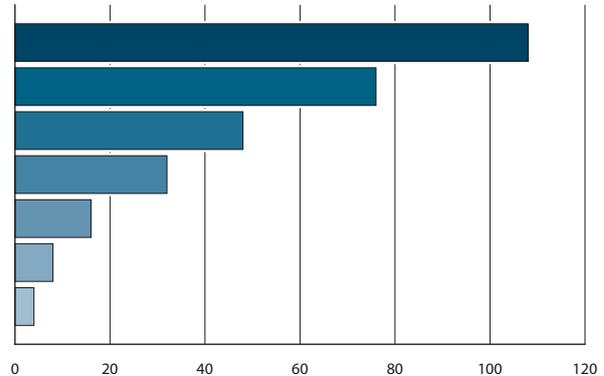
² Mackinlay, Jock (1986): „Automating the design of graphical presentations of relational information“

Im Folgenden werden die wichtigsten und gebräuchlichsten Diagrammarten dargestellt.

Balkendiagramm

- Balkendiagramme stellen den Wertebereich horizontal dar
- Eher geeignet bei langen Beschriftungen des Definitionsbereiches, wodurch auf separate Legenden verzichtet werden kann
- Sehr gut geeignet zur Darstellung von Rankings

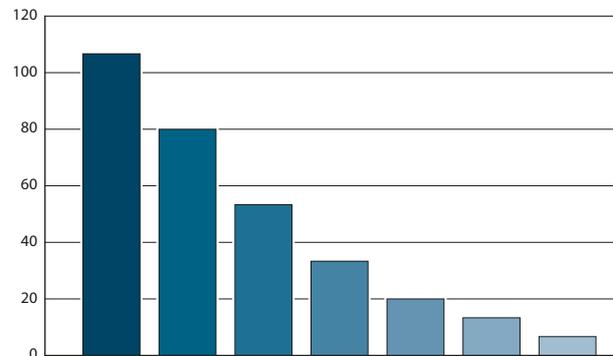
Empfohlene Datenarten ... Rangfolgen, relational
 Empfohlene Dimensionen 2D, XD



Säulendiagramm

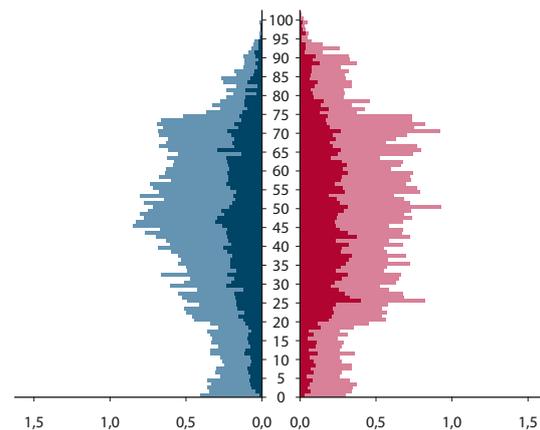
- Säulendiagramme eignen sich für ein- und zweidimensionale Vergleiche von Kategorien, Objekten, Zeiträumen
- Säulendiagramme stellen den Wertebereich senkrecht dar
- Sehr gut geeignet für die Darstellung von Schwankungen

Empfohlene Datenarten relational, temporal
 Empfohlene Dimensionen 2D, XD



Butterfly-Diagramm/Tornado Chart

- Sind eine Spezialform des Balkendiagramms
- Der Wertebereich (y-Achse) wird hier mittig angelegt, nach links und rechts werden die einzelnen Datenreihen in ihrer jeweiligen Gruppe angelegt
- Das entstehende Muster kann je nach Ausprägung wie ein Schmetterling oder eine Windhose anmuten
- Sehr gut geeignet für direkte Vergleiche zweier Gruppen mit beliebig vielen Datenreihen (häufig verwendet für Alterspyramiden nach Geschlecht)
- In Kombination mit Stapelbalken können zusätzliche Dimensionen abgebildet werden
- Häufig verwendet für Bevölkerungsdarstellungen (zum Beispiel Alterspyramide) oder Business-Case-Analyse (Risikofaktoren)

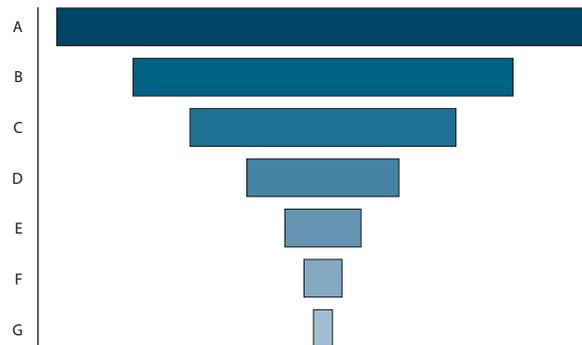


Empfohlene Datenarten Rangfolgen, relational, temporal
 Empfohlene Dimensionen 2D, XD

Trichterdiagramm

- Spezialform des Balkendiagramms, bei dem die Balken zentriert und in Form eines Trichters dargestellt werden können
- Sehr gut geeignet, um eine Rangfolge von Kategorien darzustellen
- In vielen Tools kann nur eine Datenreihe für die Darstellung verwendet werden, weshalb es für Vergleiche ungeeignet ist

Empfohlene Datenarten ... Rangfolgen, relational



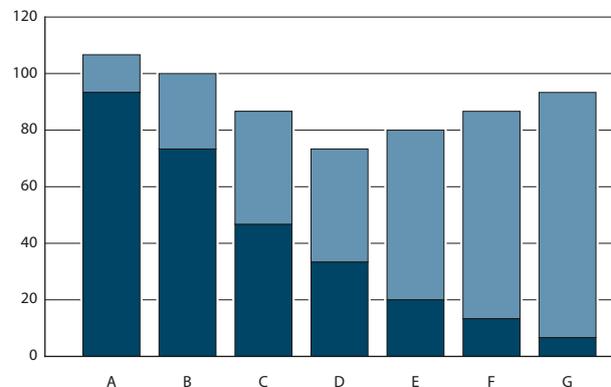
Stapeldiagramm

- Sehr gut geeignet zur gleichzeitigen Darstellung von relativen und absoluten Häufigkeiten
- Die Datenreihen werden dabei gruppiert (zum Beispiel nach Jahren) innerhalb einer Säule dargestellt, statt wie bei den meisten Säulendiagrammen nebeneinander, wodurch das Verhältnis der Einzelwerte zum Gesamtwert besser sichtbar wird

Empfohlene Datenarten

..... relational, temporal, Zusammensetzung

Empfohlene Dimensionen 2D, XD

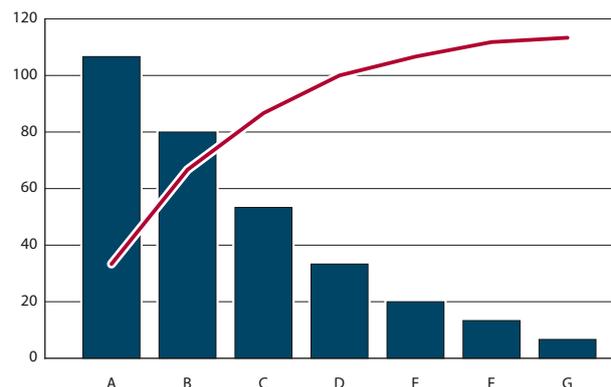


Histogramm

- Visualisiert als Spezialform des Säulendiagramms die Mengenverteilung der jeweiligen Kategorie über den Flächeninhalt der Säule
- Säulen können verschiedene Breiten haben, wenn diese Datenreihen über mehrere Kategorien abbilden
- Ähnelt von der Darstellung einem Pivot-Chart, in dem die Werte-Häufigkeiten aus Datenreihen aggregiert dargestellt werden
- Sehr häufig verwendet für die automatisierte Darstellung von Häufigkeitsverteilungen in gruppierten Kategorien, wenn eine Gruppenzuweisung in den Datensätzen nicht gegeben ist (da bei einer Histogramm-Darstellung die Gruppen frei definiert werden können)

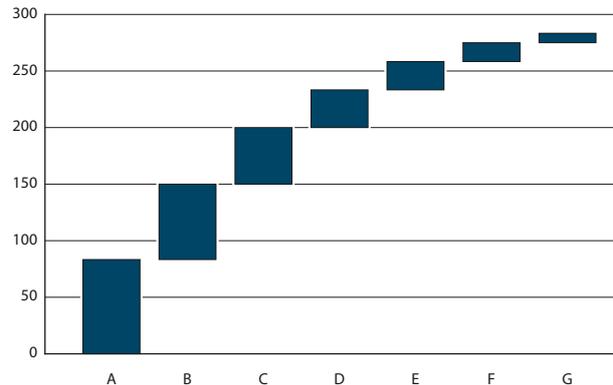
Empfohlene Datenarten Rangfolgen, relational, Zusammensetzung

Empfohlene Dimensionen 2D



Wasserfalldiagramm

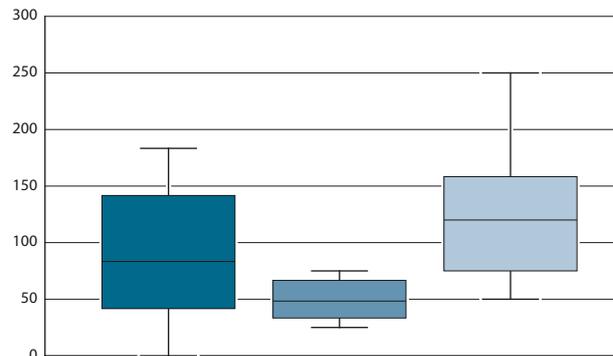
- Spezialform des Säulendiagramms, bei der ein Anfangswert durch Addition oder Subtraktion weiterer Datensäulen zu einem Endwert führt
- Sehr gut geeignet und verbreitet, um Umsatzergebnisse abzubilden
- Von Nachteil ist, dass Laien oft die Grafik nicht erschließen können, alternativ kann eine Werte-Entwicklung in einem Linien- oder Säulendiagramm dargestellt werden, bei dem nur die absoluten Werte dargestellt werden



Empfohlene Datenarten Rangfolgen, relational, Zusammensetzung
 Empfohlene Dimensionen 2D

Box-Plot-Diagramm

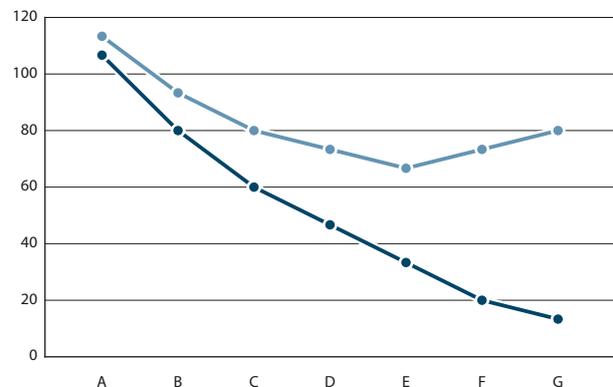
- Ein Kasten stellt horizontal oder vertikal aus einer Datenreihe gleichzeitig bis zu fünf robuste Lage- und Streuungsmaße dar (oberes und unteres Quartil beziehungsweise Eröffnungs- und Schlusskurs, Median, unterer und oberer Extremwert)
- Sehr gut geeignet für Datenexploration, Laien können die Diagramme oft nicht erschließen
- Häufig verwendet für Kurswert-Visualisierungen (zum Beispiel Aktienmarkt oder statistische Verteilungen)



Empfohlene Datenarten relational, temporal
 Empfohlene Dimensionen 2D

Linien-/Kurvendiagramm

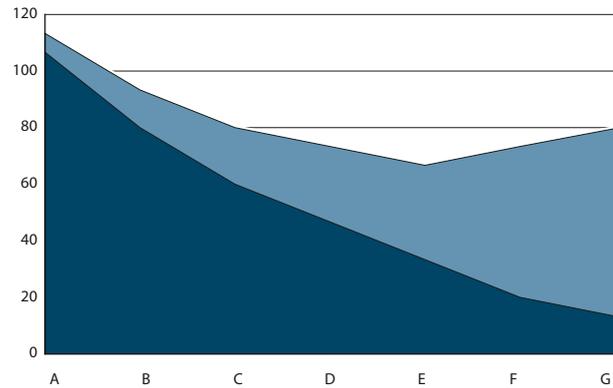
- Sehr gut geeignet für die Darstellung und den Vergleich von zeitlichen Verläufen
- Ermöglicht eine optimale Darstellung von Trendverläufen
- Verbindungslinien dienen nur der Veranschaulichung der Entwicklung zwischen den konkreten Datenpunkten
- Die x-Achse wird üblicherweise als Zeitachse verwendet



Empfohlene Datenarten relational, temporal
 Empfohlene Dimensionen 2D, XD

Flächendiagramm

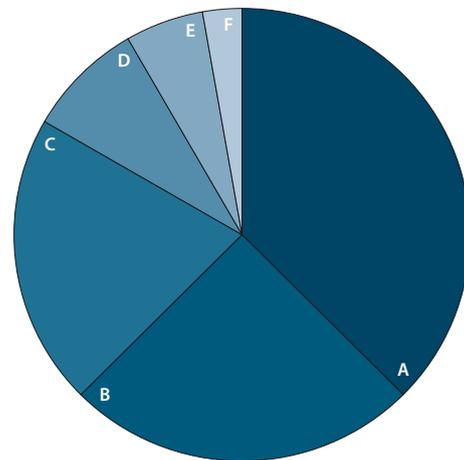
- Fläche zwischen Graphen und x-Achse ist eingefärbt und eignet sich bei Vergleichen sehr gut für das Hervorheben von Unterschieden
- Sehr gut geeignet für die Darstellung und den Vergleich von zeitlichen Verläufen
- Ermöglicht eine optimale Darstellung von Trendverläufen
- Verbindungslinien dienen nur der Veranschaulichung der Entwicklung zwischen den konkreten Datenpunkten
- Die x-Achse wird üblicherweise als Zeitachse verwendet



Empfohlene Datenarten relational, temporal, Zusammensetzung
 Empfohlene Dimensionen 2D, XD

Kreis- oder Ringdiagramm

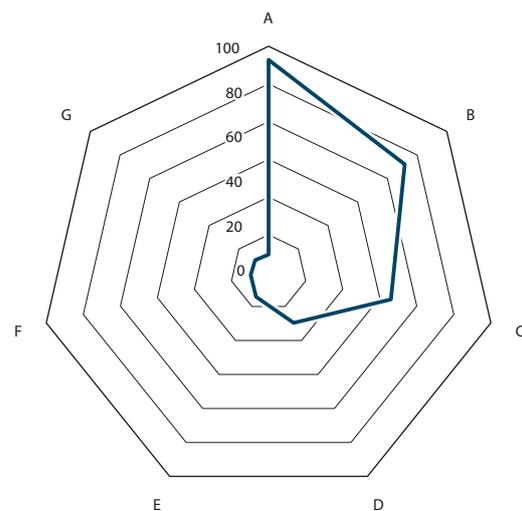
- Datenreihen werden als „Tortenstücke“ dargestellt und visualisieren damit sehr gut das Verhältnis der Datenreihen untereinander und zur Gesamtheit
- Nicht geeignet für die Abbildung zeitlicher Verläufe



Empfohlene Datenarten Zusammensetzung
 Empfohlene Dimensionen 2D

Netz-/Sterndiagramm

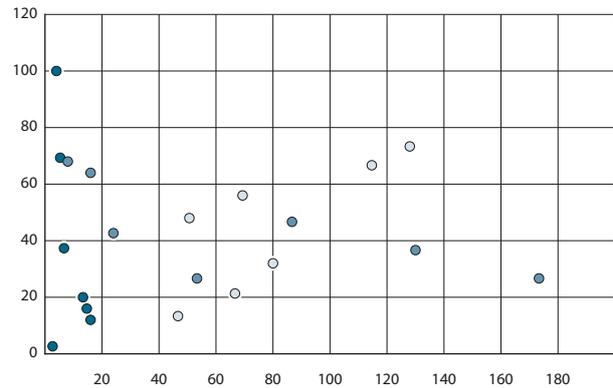
- Stellt die Ausprägung bestimmter Kategorien/ Datenreihen an beliebig vielen, gleichwinklig verteilten Achsen innerhalb eines Kreises dar
- Durch die entstehende Linie der verbundenen Datenpunkte lassen sich Profile erstellen, die sich sehr gut für Vergleiche von Objekten eignen
- Wird oft für Benchmarking von Firmen verwendet
- Ein Nachteil ist, dass die Grafik von Laien oft nicht auf Anhieb erschlossen werden kann
- Alternativ können die einzelnen Definitionsbereiche mit separaten Diagrammen entzerrt werden, die schneller zu erschließen sind



Empfohlene Datenarten relational, Zusammensetzung
 Empfohlene Dimensionen 2D, XD

Streudiagramm (Scatter Plot)

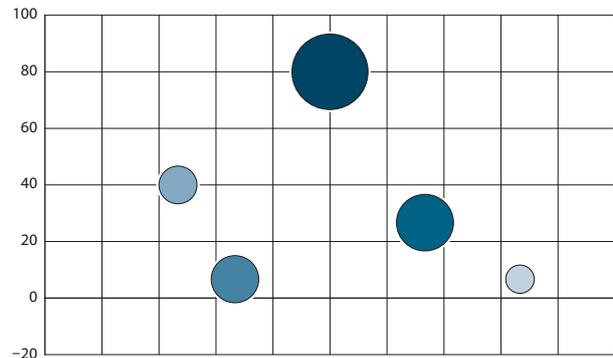
- Datenreihen werden in Form von Punkten visualisiert
- Sehr gut geeignet zur Darstellung von Häufigkeitsverteilungen und für viele zusammenliegende Werte
- Durch viele gehäufte Werte entsteht eine Punktwolke, die sehr gut große Mengen visualisiert
- Häufig verwendet für zum Beispiel Emissions-Visualisierung



Empfohlene Datenarten relational, temporal
 Empfohlene Dimensionen 2D, XD

Blasen-/Portfoliodiagramm

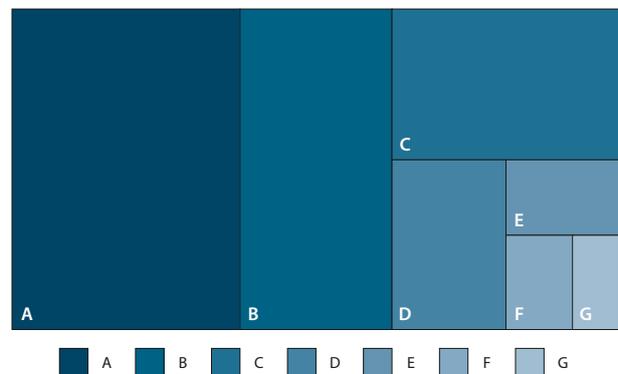
- Das Blasen- und Portfoliodiagramm kann als Weiterentwicklung des Punktdiagramms betrachtet werden, bei dem in Bezug auf das Diagramm die Datenpunkte in der Größe variabel sind und damit eine weitere Dimension abbilden
- Sehr gut geeignet für die Darstellung multivariater Daten
- Portfoliodiagramme erweitern das Blasen- und Portfoliodiagramm um eine Einteilung der Diagrammfläche in verschiedene Bereiche/ Kästen
- Die Kästen dienen der schnelleren Zuordnung zu bestimmten Kategorien



Empfohlene Datenarten relational, temporal
 Empfohlene Dimensionen 2D, XD

Treemap

- Treemaps dienen besonders der Visualisierung von Hierarchien
- Die Verzweigungen werden durch Flächen- und farbliche Zuordnungen visualisiert
- Sehr gut geeignet für den Vergleich von Mengenverhältnissen bei gleichzeitigem Fokus auf die beinhaltenden Kategorien
- Für Vergleiche von Mengenverhältnissen über verschiedene Ebenen eignet sich jedoch eher das Sunburst-Diagramm, das jedoch die Kategorien nicht derartig übersichtlich darstellt



Empfohlene Datenarten Graphdaten, Rangfolgen, Zusammensetzung
 Empfohlene Dimensionen 2D, XD

2.6. Leitlinie „Datenschutz und Datensicherheit“

Ziel ist die Nutzbarmachung von Daten der städtischen Fachbereiche auf einem Informationsportal für die Öffentlichkeit mittels eines Data Warehouse (DWH), in welchem neben der Informationsbereitstellung auch Datenanalysen vorgenommen werden können. Die Daten sollen dort lesbar dargestellt und erfahrbar aufbereitet werden können, sodass Datenanalysen und -interpretationen durchführbar sind.

Der Begriff Informationsportal meint im Sprachgebrauch dieses Leitfadens eine Softwareplattform, auf der mittels unterschiedlicher Architekturen, Technologien und Produkte bereits erhobene kommunale Daten eingebunden beziehungsweise gespeichert, verarbeitet und zu Analyse- und Auswertungszwecken bereitgestellt werden. Der Leitfaden „Datenschutz und Datensicherheit“ soll für die Beschaffung und Einführung einer solchen Lösung Handlungsempfehlungen geben. Diese resultieren aus der Einhaltung der Anforderungen an Datenschutz und Datensicherheit und folgen dem Prozess der Einführung eines solchen Portals von der Konzeption bis zur Produktivsetzung.

Im Rahmen der Erstellung dieses Leitfadens wurde der Datenschutz- und Informationssicherheitsbeauftragte der Stadt Cottbus/Chósebus befragt. Seine Aussagen sind in die folgenden Abschnitte des vorliegenden Leitfadens integriert.

Datensicherheit ist eine Voraussetzung für den Datenschutz und stellt einen Datenschutzgrundsatz dar. Damit personenbezogene Daten geschützt sind, müssen entsprechende technische Maßnahmen für ihre Sicherheit etabliert werden. Datenschutzmaßnahmen sorgen des Weiteren dafür, dass personenbezogene Daten nicht nur technisch geschützt und sicher verwahrt sind, sondern es sollen auch ihre Erhebung, Verarbeitung und Weitergabe eingeschränkt werden, um die Privatsphäre der betreffenden Personen zu schützen.

2.6.1. Rechtsgrundlagen eines kommunalen Informationsportals

Ausgangspunkt für die datenschutzrechtliche Betrachtung des Informationsportals sind die gesetzlichen Grundlagen, welche die Verarbeitung von personenbezogenen Daten regeln. Dazu gehört das sogenannte allgemeine Datenschutzrecht mit folgenden Normen:

- Datenschutz-Grundverordnung (DSGVO)
- Bundesdatenschutzgesetz (BDSG)

Für die Stadt Cottbus/Chósebus waren darüber hinaus folgende Gesetze und Leitlinien zu berücksichtigen:

- Brandenburgisches Datenschutzgesetz (BbgDSG)

Leitlinien zur Informationssicherheit der Stadt Cottbus/Chósebus i. V. m. BSI-Standard 100-1 und 100, normiert durch Gesetz über die elektronische Verwaltung im Land Brandenburg (Brandenburgisches E-Government-Gesetz (BbgEGovG))

Zudem gelten besondere Regelungen wie zum Beispiel das Brandenburgische Polizei-, Justizvollzugs- und Maßregelvollzugsdatenschutzgesetz. Ebenfalls sind spezielle datenschutzrechtliche Regelungen in den Fachrechten enthalten, beispielsweise in der Abgabenordnung oder dem Sozialgesetzbuch. Für IT-Anwendungen gelten die Vorgaben für den IT-Grundsatz des Bundesamtes für Sicherheit in der Informationstechnologie (BSI).

Die Dokumentation der datenschutz- und informationssicherheitsbezogenen Ziele und entsprechender Maßnahmen von Beginn an ist sinnvoll. Dies gilt nicht nur im Einzelnen in Bezug auf dieses Verfahren, sondern um-

fasst grundsätzlich, die Empfehlungen, Ziele, Maßnahmen und Prozesse für Verfahren zu dokumentieren. Zum einen wird dadurch das im Sinne der Datenschutz-Grundverordnung geforderte Prüfmittel für Prüfinstanzen erzeugt, zum anderen eröffnet die Dokumentation die Möglichkeit, die Prozesse zu klassifizieren und hinsichtlich der Qualität, Effizienz und Medienbrüche zu bewerten. Nur auf dieser Grundlage ist eine anschließende kontinuierliche Prozessverbesserung möglich. In den folgenden Abschnitten werden Hinweise zur Umsetzung eines datenschutzkonformen Informationsportals gegeben.

2.6.2. Informationsportal – Einführungsprozess und erforderliche Maßnahmen

Zu beachten ist, dass bereits vor Beginn des Einführungsprozesses eine Reihe von Maßnahmen definiert wird bzw. wurde, mit deren Hilfe während des gesamten Prozesses die Einhaltung des Datenschutzes gewährleistet werden kann. Für die einzelnen Phasen der Einführung werden im Folgenden Maßnahmen aufgeführt, welche diese Forderung unterstützen.

Spätestens mit dem Beginn der Überlegungen für das vorgesehene Informationsportal sind geeignete Data Governance-Strukturen für den Umgang mit Daten und die Auswertungsinteressen selbst festzulegen. Diese Regeln umfassen konkrete Prozesse, Standards und Richtlinien in Bezug auf die Datenerhebung, -verarbeitung und -analyse sowie die Datenbereitstellung in der Verwaltung. Sie definieren Rollen mit dazugehörigen Aufgaben und Verantwortlichkeiten, die unter anderem die Behandlung datenschutzrechtlicher Aspekte betreffen.

Data Governance-Strukturen

- Definieren einen einheitlichen Begriffsrahmen für den Umgang mit Daten
- Geben Prozesse, Verfahren und Standards von der Formulierung des Auswertungsinteresses bis zur Bereitstellung von Analysen vor
- Definieren Rollen und Kompetenzen, die für den sachgerechten Umgang mit Daten notwendig sind
- Definieren datenbezogene Merkmale und -prozesse, welche die Validität und Qualität der Daten sichern (Compliance-Regelungen)

2.6.2.1. Antragstellung (bei der für die Beschaffung zuständigen Stelle)

Der Zweck der Antragstellung ist, Klarheit darüber zu erlangen, ob mithilfe des geplanten Informationsportals personenbezogene Informationen in Form von IP-Adressen (Adresse in Computernetzen, die – wie das Internet – auf dem Internet Protocol (IP) basiert), Nutzerdaten oder personenbezogenen Daten verarbeitet werden sollen. Der Datenschutz- und Informationssicherheitsbeauftragte hat eine beratende Funktion und nimmt in einem Umsetzungsvorhaben eine Richtungs- und Kontrollpflicht wahr.

Antragstellung

- Die Ziele, die mithilfe des Informationsportals erreicht werden sollen, sind im Projektantrag zu dokumentieren und nachvollziehbar zu begründen. Dies schafft die Grundlage für weitere Prüfmaßnahmen.
- Der Datenschutz- und Informationssicherheitsbeauftragte soll in die Antragstellung des Vorhabens miteinbezogen werden. Hierbei kann eine Vorprüfung des Antragsentwurfes erfolgen, der frühzeitig Schwächen und Verbesserungsbedarfe aufzeigt.
- Ebenfalls sind die Datenverwalter der beteiligten Fachbereiche miteinzubeziehen. Sie haben später den Auftrag, die Anforderungen und Maßnahmen für Datenschutz und Informationssicherheit zu erarbeiten.

- Definieren datenbezogener Merkmale und Prozesse, welche die Validität und Qualität der Daten sichern (Compliance-Regelungen).

2.6.2.2. Konzeptions- und Planungsphase

Am Beginn einer Softwareeinführung steht eine Konzeptions- und Planungsphase, in der fachliche und nicht-fachliche Anforderungen an das Informationsportal sowie die organisatorischen Prozesse im Zusammenhang mit der Datenverarbeitung erhoben und dokumentiert werden. Vorgaben entsprechend den Anforderungen des BSI-Grundschutzkataloges sowie der DSGVO (unter anderem sind Softwareprodukte vorzusehen, die durch ihre technische Gestaltung und Voreinstellungen als datenschutzfreundlich gelten) sind zu berücksichtigen. Diese Vorgaben sind bereits bei der Konzeption einer Software beziehungsweise beim Entwurf des zukünftigen Systems zu beachten, denn sie gelten sowohl für die Datenverarbeitungsprozesse innerhalb des Systems als auch für die Datenspeicherung. Diese Anforderungen sind zwingend zu dokumentieren.

Datenanalytische Anforderungen

Bereits während der Konzeptionsphase sind die analytischen Auswertungsinteressen an die Daten zu formulieren. Ist die Auswertung personenbezogener Daten von Interesse, kann eine Funktion des Informationsportals zur Anonymisierung beziehungsweise Pseudonymisierung von Daten helfen, die Datenschutz-Grundverordnung-Konformität zu erreichen und für Auswertungszwecke nutzbar zu machen. Oft ist eine Datenanalyse auf der Basis von personenbezogenen Daten nicht notwendig, sondern kann ebenso über die Aggregation von Merkmalen (zum Beispiel Wohnort, Geschlecht, Alter) erfolgen und so eine ausreichende Auswertungsbasis schaffen.

Datenschutz

- Welche Datenbasis ist von den analytischen Interessen betroffen? Bestehen hierfür bereits Regelungen und Schutzanforderungen?
- Werden beziehungsweise müssen personenbezogene Daten verarbeitet werden?
- Werden durch die Kombination von (personenbezogenen) Daten neue Daten generiert, welche ebenfalls Merkmale des Personenbezuges aufweisen?
- Können personenbezogene Daten nach der Verarbeitung (Pseudonymisierung) nachweislich gelöscht werden?

Sicherheitstechnische Anforderungen

Es muss möglich sein, die Daten des Informationsportals sicher und vertraulich zu speichern. Dies bedeutet, dass eine dem Stand der Technik entsprechende Datenverschlüsselung bei der Speicherung und Übertragung sowie eine Ende-zu-Ende-Verschlüsselung und Mechanismen zur Sicherstellung von Integrität und Authentizität der Daten vorhanden sein müssen. Des Weiteren sind, den IT-Grundschutz-Standards des BSI folgend, weitere Anforderungen zur Einhaltung der Datensicherheit zu erheben und in die Konzeption mit einfließen zu lassen.

Datensicherheit

- Die Anwendung ist durch geeignete Maßnahmen gegen unbefugten Zugriff geschützt.
- Die Informationsweiterleitung an andere Verfahren über Schnittstellen ist dokumentiert.
- Ein Prozess zum Umgang mit auftretenden Sicherheitslücken ist definiert. Es stehen entsprechende Verfahren zum Update/zur Beseitigung der Sicherheitslücken bereit.

- Datensicherungen werden durchgeführt und Wiederherstellungsprozesse definiert.
- Zugriffe auf das Informationsportal werden datenschutzkonform protokolliert.

2.6.2.3. Implementierungs- und Testphase des kommunalen Informationsportals

Die Umsetzung von angepassten oder Individuallösungen mittels agiler Methoden oder die Beschaffung und Implementierung von Standardsystemen bedeutet unabhängig vom Vorgehensmodell, dass die definierten Anforderungen geprüft und bestätigt werden. Erst dann kann eine Nutzung im Produktivbetrieb erfolgen. Die Prüfung der Implementierung erfolgt mittels Verfahrenstests. Dabei muss die gewählte Lösung die anforderungsgemäße Funktionalität aufweisen. Zudem ist die Wirksamkeit der im Sicherheitskonzept festgelegten Schritte nachzuweisen. Um den Anforderungen an den Datenschutz zu entsprechen, sind die Prüfverfahren und -ergebnisse sowie die für die Testmethoden eingesetzten Werkzeuge zu protokollieren.

Die Schwierigkeit bei der Prüfung von datenbasierten Verfahren liegt häufig in der Bereitstellung geeigneter Testdaten. Gerade wenn es sich um eine Massendatenverarbeitung handelt und sich das System noch in der Entwicklung befindet, ist eine Funktionsprüfung mit Echtdateien nicht zulässig. Um diesem Umstand frühzeitig Rechnung zu tragen, sind in der Projektplanung daher entsprechende Schritte und Maßnahmen zu planen, die eine rechtzeitige Beschaffung und Bereitstellung von geeigneten Testdaten vorsehen. Auch ein sogenannter Probetrieb in Pilotbereichen kann nur dann erfolgen, wenn die Wirksamkeit der konzipierten Datenschutzmaßnahmen geprüft und bestätigt wurde.

Testphase

- Die Prüfung der konzipierten Datenschutz- und Datensicherheitsmaßnahmen durch Tests muss nachgewiesen werden. Der Datenschutz- und Informationssicherheitsbeauftragte oder die Datenschutzkoordinatoren der Fachbereiche können bei der Testdurchführung beziehungsweise der Bewertung der jeweiligen Funktionalitäten unterstützen.
- Für die Funktionsprüfung notwendige Testdaten stehen zur Verfügung. Dabei sind auch Negativtests durchzuführen, das heißt Kontrollen, die nachweisen, dass eine fehlerhafte Bedienung oder provozierte Verstöße gegen Maßnahmen des Datenschutzes oder der Informationssicherheit kein unerwünschtes Systemverhalten zur Folge haben.
- Die zukünftigen Nutzenden werden frühzeitig geschult und mit den datensensiblen Bereichen und Prozessen vertraut gemacht.

2.6.2.4. Begleitung und Überwachung der Wirkbetriebsphase

Wie in der Konzeptionsphase beschrieben, muss der Wirkbetrieb regelhaft ablaufen. Hinsichtlich der Schaffung von Transparenz und der Dokumentation von Prozessen muss daher regelmäßig geprüft werden,

- ob die Verfahrensdokumentation noch den tatsächlichen Abläufen entspricht,
- ob Änderungsanforderungen zu neuen Funktionen geführt haben, deren Auswirkung auf die etablierten Schutzmaßnahmen geprüft und bestätigt werden muss, oder
- ob die Lösung noch den aktuellen gesetzlichen Vorgaben entspricht.

Aktivitäten von Nutzenden und Administratoren sind in dem Maße zu protokollieren, wie es dem Zweck der Datensicherheit dient (und nicht zur personenbezogenen Verhaltenskontrolle). Stichprobenhaft ist eine Überprüfung der Regeleinholung in der Datenverarbeitung durchzuführen.

Wirkbetrieb

- Funktionen, die den Datenschutz und die Informationssicherheit der Anwendung betreffen, werden regelmäßig überprüft.
- Werden Softwareverbesserungen, -aktualisierungen oder der Funktionsumfang des Programmes erweitert beziehungsweise verbessert, sind die Funktionsfähigkeit und der anforderungsgemäße Betrieb nachzuweisen.
- Bei funktionalen Erweiterungen ist ebenfalls durch Tests die Einhaltung der Regelungen von Datenschutz und Informationssicherheit zu belegen.
- Für das System sind zyklisch Audits, die untersuchen, ob Prozesse, Anforderungen und Richtlinien die geforderten Standards erfüllen, durchzuführen.

2.6.3. Datensicherheit und Nutzung von Cloud-Diensten

Wird in der Konzeptionsphase festgelegt, dass das Informationsportal oder dessen Komponenten durch externe Anbieter außerhalb der eigenen Infrastruktur der Stadt Cottbus/Chósebus in Form von Cloud-Dienstleistungen erbracht werden sollen, ist diese externe Datenverarbeitung vergaberechtlich mit einem definierten Leistungskonzept auszuschreiben. Für die datenschutzkonforme Nutzung von Software-as-a-Service-Dienstleistern (SaaS) gelten folgende Anforderungen:

- Der rechtliche Rahmen der Dienstenutzung wird durch die Datenschutz-Grundverordnung bestimmt.
- Der Anbieter muss seine Dienste im Geltungsbereich der Datenschutz-Grundverordnung anbieten, das heißt die Datenverarbeitung muss in der Europäischen Union erfolgen.
- Art. 48 der DSGVO verbietet Unternehmen zudem die Übergabe von in der EU gesicherten Daten ohne Rechtshilfeabkommen.

Die Nutzung von Anbietern, die ihre Dienste außerhalb des Geltungsbereiches der Datenschutz-Grundverordnung anbieten, erfordert eine gesonderte Überprüfung. So unterliegen beispielsweise Unternehmen mit Sitz in den Vereinigten Staaten von Amerika (USA) dem Clarifying Lawful Overseas Use of Data Act or the CLOUD Act (Cloud Act) und müssen gemäß dieser Norm auf Verlangen sind in ihrer Obhut liegende Daten den US-Regierungsbehörden zur Verfügung stellen, auch wenn diese außerhalb der USA gespeichert sind. Der CLOUD Act ist nicht nur auf in den USA ansässige Unternehmen und Dienstleister beschränkt. Er gilt auch dann, wenn eine Niederlassung oder eine Geschäftstätigkeit in den USA vorliegen. Öffentliche Nutzende (zum Beispiel Kommunen) verstoßen demzufolge gegen die Datenschutz-Grundverordnung, wenn Cloud-Services von US-Anbietern genutzt sowie personenbezogene Daten mithilfe solcher Dienstleister gespeichert und verarbeitet werden.

Die beispielhafte Betrachtung des CLOUD Act ist vor allem in die Überlegung miteinzubeziehen, wenn große Diensteanbieter wie Microsoft, Oracle oder Google genutzt werden. Zu beachten ist grundsätzlich, dass auch Anbieter aus anderen Staaten außerhalb der Europäischen Union als Diensteanbieter auftreten. Hier gilt ebenfalls der Grundsatz, dass eine Dienstleistung im Geltungsbereich der Datenschutz-Grundverordnung stattfinden muss. Hinsichtlich der Beurteilung der Informationssicherheit von SaaS-Diensten können die „C5 Kriterien zur Beurteilung der Informationssicherheit von Cloud-Diensten“ des Bundesamtes für Sicherheit in der Informationstechnik (Cloud Computing Compliance Criteria Catalogue) herangezogen werden. Entsprechende Prüfnachweise sind durch den Betreiber des Cloud-Dienstes in Form von Zertifikaten oder Audit-Nachweisen zu erbringen.

Nutzung von SaaS-Diensten

- Die Speicherung und Verarbeitung erfolgt innerhalb des Geltungsbereiches der Datenschutz-Grundverordnung (Europäische Union). Der Anbieter ist nicht Teil eines US-Konzerns oder seiner Tochtergesellschaften.
- Es wird eine Auftragsdatenverarbeitungsvereinbarung abgeschlossen.
- Zertifizierungen des Anbieters geben darüber Auskunft, ob Datenschutzstandards eingehalten werden (zum Beispiel ISO27001, EN 50600, Eco Datacenter Star Audit, TÜViT) und eine regelmäßige und unabhängige Auditierung erfolgt.
- Die Lösung bietet Rechte- und Rollenmanagement und unterstützt Kunden bei Fragen des Datenschutzes und der Informationssicherheit.
- Der Dienstanbieter kann vertraglich verpflichtet werden, den Nutzenden über eine Expansion in die USA und damit in den Gültigkeitsbereich des CLOUD Act zu informieren.

3. Systemarchitektur für das Datenmanagement

Um die Implementierung des Datenmanagements zu vertretbaren operativen Kosten sowie die Administration der komplexen Zusammenhänge von Daten zu ermöglichen, ist die technische Verzahnung der prozessualen Ausgestaltung des Datenmanagements mit der bestehenden beziehungsweise der gewünschten IT-Systemarchitektur unabdingbar. Ein effizientes und innovatives Datenmanagement benötigt einen hohen Anteil an IT-Unterstützung, um sowohl die Prozesse der Datenaufbereitung und -verteilung als auch das Datenvolumen zu bewältigen. Für eine erfolgreiche Umsetzung sind gegebenenfalls bestehende Rahmenbedingungen anzupassen und sicherzustellen, dass Fachbereiche und IT-Verantwortliche sich proaktiv abstimmen und sich mit den fachlichen Anforderungen an das Datenmanagement auseinandersetzen.

Am Modellprojekt sind seitens der Stadt Cottbus/Chósebus drei Fachbereiche (Fachbereich 33 Statistikstelle, Fachbereich 61 Stadtentwicklung und Fachbereich 62 Geoinformation und Liegenschaftskataster) sowie als Leiter des kommunalen Forschungsvorhabens der Chief Information Officer (CIO) der Stadt Cottbus/Chósebus beteiligt. Mit einzelnen Personen aus den Fachbereichen Stadtentwicklung, Geoinformation und Liegenschaftskataster sowie dem Team Statistik der Stadtverwaltung wurden halbstrukturierte Interviews, Workshops sowie eine Abfrage mittels Online-Fragebogen hinsichtlich der Anforderungen an ein Informationsportal (mit Schwerpunkt auf die Funktionalitäten der Datenanalytik) durchgeführt. Ziele der Gespräche mit den Vertretern der beteiligten Fachbereiche waren

- eine Sachstandsabfrage zum Einsatz analytischer Methodik im Arbeitsalltag,
- die Abfrage des Bedarfs an datenanalytischen Auswertungen im Kontext der fachlichen Aufgabeninhalte sowie
- die Erhebung von Anforderungen an ein Informationsportal hinsichtlich der benötigten datenanalytischen Auswertungen.

3.1. Bedarf und Anforderung in der kommunalen Verwaltung

Die folgenden datenanalytischen Funktionalitäten werden von den befragten Vertretern der Fachbereiche der Stadt Cottbus/Chósebus für die tägliche Arbeit als bedeutend angesehen, die momentan nicht oder eingeschränkt mit IT-Unterstützung durchzuführen sind:

- Möglichkeit der kartografischen Darstellung (Fachbereich 33 Statistikstelle, Fachbereich 61 Stadtentwicklung)
- Verknüpfungen/Verschneidung von Daten aus anderen Fachbereichen (Fachbereich 61 Stadtentwicklung)
- Verschneidung von Daten, Analyse zur Datenaktualität, Verknüpfung verschiedener Datenquellen, Ergebnispräsentation in verständlicher Form, beispielsweise als Dashboard mit verschiedenen Diagrammtypen (Fachbereich 62 Geoinformation und Liegenschaftskataster)
- Möglichkeit zur Reporterstellung (alle Befragten)
- Grafischer Anspruch, Verfügbarkeit der Interpretationen von Analysen (Fachbereich 61 Stadtentwicklung)

- Anpassung der Eingangsdefinitionen, ohne die Analyse komplett neu zu beginnen, aber Änderungen leicht erkennen zu können (zum Beispiel „Was passiert, wenn der Umkreis um 100 m vergrößert wird“), Individualisierbarkeit von Auswertungen (Fachbereich 62 Geoinformation und Liegenschaftskataster)

Einem zusätzlichen Analytik-Tool, welches neben der bereits eingesetzten Software zur Anwendung kommen könnte, sprechen zwei von fünf Befragten der Online-Umfrage keinen Mehrwert zu, obwohl die Teilnehmenden andererseits von mangelhafter Nutzung von datenanalytischen Funktionalitäten in ihrem Tätigkeitskontext sprechen. Die Teilnehmenden aus dem Fachbereich 62 Geoinformation und Liegenschaftskataster, die keinen Mehrwert in einem zusätzlichen Tracking-Tool sehen, begründen ihre Aussagen damit, dass einem zusätzlichen Analytik-Tool zwar insgesamt ein hoher Wert zugesprochen wird, momentan jedoch die bereits zu nutzende Software ausreichend ist und im Kontext der GIS-Nutzung keine weitere Anwendung notwendig sei.

Beschäftigte aus dem Fachbereich 62 Geoinformation und Liegenschaftskataster gehen von einem Mehrwert hinsichtlich datenbankübergreifender Datenanalyse und -suche aus. Die Teilnehmenden des Fachbereiches 33 Statistikstelle weisen einem zusätzlichen Tool einen hohen Mehrwert zu, da damit vorhandene Informationen gezielt abgefragt werden können. Folgende Vorteile ergeben sich:

- eine Einsparung von Arbeits- und Zeitaufwand,
- es wird von einer Verringerung der Fehlerquote (zum Beispiel durch Plausibilitätsprüfungen) ausgegangen,
- mehr/bessere Auswertungsmöglichkeiten geboten werden.

Die Auswertung des Fragebogens lässt außerdem darauf schließen, dass Qualifizierungsbedarf (beispielsweise Schulungen, Weiterbildungen oder Trainings) im Hinblick auf die Nutzung und Anwendung datenanalytischer und statistischer Methoden besteht. Momentan häufig genutzte Funktionen beschränken sich beispielsweise auf SUMME, MITTELWERT, MIN, MAX, Vergleiche oder WENN-DANN-SONST, was auf eine eher geringe Nutzung weiterer statistischer Methoden hinweist.

Zudem wurde angegeben, dass bereits vorhandene Statistiktools aufgrund ihrer Komplexität ungenutzt bleiben. Die Fachanwendenden äußern hier Qualifizierungsbedarf. Außerdem stellte sich in den Workshops und Befragungen heraus, dass es sehr unterschiedliche Vorstellungen seitens der Anwendenden bezüglich des Begriffes Statistik gibt. Während einige Anwendende einzelne statistische Funktionen bereits nutzen oder in Zukunft nutzen möchten, definieren andere Statistik als Zugriffstatistiken auf Daten. Dennoch sollte die für alle Fachanwendenden erforderliche Befähigung im vielschichtigen Gebiet der Statistik erlangt werden, damit diese die berechneten Daten korrekt interpretieren beziehungsweise weitere Anwendungsmöglichkeiten für ihre Daten erkennen können. Dies ist erst dann sinnvoll, wenn ein konkretes Instrument zur Verfügung steht, anhand dessen die entsprechenden Schulungen durchgeführt werden können.

3.2. Data Warehousing und Datenbanksysteme

Auf der Basis der erhobenen Anforderungen und Erkenntnisse hat die Andreas Krohn & Birk Heinze GbR eine Marktbeurteilung vorgenommen und geeignete Softwareprodukte untersucht. Im Fokus standen dabei Lösungen, die rechtssicher betrieben werden können, idealerweise On-Premise im eigenen kommunalen Rechenzentrum.

Durchgängige Open Source-Lösungen oder DSGVO-konforme Lösungen deutscher beziehungsweise europäischer Hersteller und Betreiber sind hier bevorzugt auszuwählen. Anhand dieser Kriterien wurden viele reine

Cloud-Lösungen US-amerikanischer Hersteller zeitig aus der Betrachtung ausgeschlossen (siehe hierzu auch Kapitel „Datensicherheit und Nutzung von Cloud-Diensten“).

3.2.1. DWH mit Datavault Builder™

Ein Datavault Builder™ unterstützt die Anbindung zahlreicher heterogener Datenquellen. Mit seiner browserbasierten, intuitiven Nutzeroberfläche ermöglicht er die einfache Bereitstellung der Rohdaten durch den Fachanwendenden. Dabei wird sowohl die Bottom-Up- als auch die Top-Down-Vorgehensweise vollständig unterstützt. In kurzer Zeit können somit die ersten Schichten nach Data Vault aufgebaut werden. Diese bilden die Ausgangsbasis für die Entwicklung der ersten BI-Reports mit der favorisierten BI-Lösung. Mit dem Datavault Builder™ können zahlreiche Aufwände minimiert werden, da insbesondere der Entwicklungsprozess vollständig toolgestützt und visuell durchlaufen wird. Durch die Entkopplung der bestehenden Datenobjekte beziehungsweise Entitäten im Modell ermöglicht die Data Vault-Methodik ein hochgradig paralleles Beladen und damit auch eine effiziente Weiterverarbeitung der Daten bis in die Reporting-Schicht des Data Warehouse. Daten werden schnell aktualisiert und können rasch zur Verfügung gestellt werden. Besonders intuitiv sind hier die grafische Modellierungskomponente und die Data Lineage-Visualisierung.

Tabelle 11: Vor- und Nachteile Datavault Builder™ (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Sehr hoher Reifegrad	Höherer initialer Aufwand durch Erstellung des zentralen Data Warehouse im Vergleich zu einer Self Service BI-Lösung mit integrierter Datenhaltung
Komplexe Modellierung und Integration möglich	
Open-Source DBMS PostgreSQL wird unterstützt	
Exasol DB als selbstwartendes High Performance DBMS wird unterstützt	
Gängige BI-Tools setzen auf DWH auf	
Volle DWH-Funktionalität nach Data Vault	

3.2.2. DWH mit Qlik Compose

Die Qlik Plattform stellt mit Qlik Compose ebenfalls ein Werkzeug zur Data Warehouse-Automatisierung basierend auf dem Data Vault-Ansatz zur Verfügung. Das Data Vault-Datenmodell wird hier aber im Hintergrund gehalten. Intern werden Hub- und Satelliten-Tabellen für die Entities (Geschäftsobjekte) und deren Attribute erzeugt. Die Historisierung findet gemäß Data Vault in den Satelliten-Tabellen statt. Die Anwendenden selbst sehen im Modell Entities, Attribute, Attribut Domains und Relationen. Diese können aus den Datenquellen extrahiert, von einem externen erwin-Modell geladen oder manuell erstellt werden. Die Lösung automatisiert das Warehouse-Design, das Erzeugen des ETL-Codes für die Transformation der Daten in die einzelnen DWH-Layer und das Durchführen der Datenaktualisierungen. Die erzeugten Data Marts bilden die Basis für die BI-Reports und -analysen. Dabei können sowohl QLIK Sense als auch alternative BI-Lösungen verwendet werden.

Die Plattform ermöglicht über eine browserbasierte Nutzeroberfläche einen intuitiv gestalteten Self Service-Zugang. Neben Data Warehouse-Projekten werden auch Data Lake-Projekte unterstützt.

Tabelle 12: Vor- und Nachteile Qlik Compose (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Hoher Reifegrad	Höherer initialer Aufwand durch Erstellung des zentralen DWH im Vergleich zu einer Self Service BI-Lösung mit integrierter Datenhaltung
Komplexe Modellierung und Integration möglich	Sehr teure Katalog-Funktionalität
Integration und BI aus einer Hand	PostgreSQL und EXASOL werden nicht als DWH DBMS unterstützt
Gängige BI-Tools setzen auf DWH auf	
Volle DWH-Funktionalität nach Data Vault	

3.2.3. Self-Service Analytics mit Qlik Sense

Qlik Sense ist eine vollständige Data Analytics-Plattform, die über eine assoziative Analyse-Engine, und eine leistungsfähige Cloud-Plattform verfügt und Künstliche Intelligenz nutzt.

Tabelle 13: Vor- und Nachteile Qlik Sense (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Sehr hoher Self Service-Grad	Ein höherer Schulungsaufwand für die Nutzenden ist erforderlich
Kurze Entwicklungszeiten	Kein komplexes Modelling (zentrales Datenmodell)
Ad-hoc-Abfragen durch assoziativen Ansatz möglich	Historisierung nur bedingt möglich (manuell)
Geringerer Administrationssaufwand	Entkopplung der operativen Systeme über Zwischenschicht (zum Beispiel Landing Zone) nötig
Gute Reporting-Funktionen	Sehr teure Katalog-Funktionalität
	Export nur über QVD-Files oder manuell in CSV/Excel

3.2.4. Tableau Plattform (DWH light)

Die Bereitstellung der Analysen und der Erkenntnisse sollen nicht nur einzelnen Beschäftigten, sondern der gesamten Verwaltung dienen. Bei dieser Applikation handelt es sich um eine sichere und flexible End-to-End-Analytics-Plattform (durchgängig von den Datenquellen bis zu den Endergebnissen wie Reports oder Dashboards), die unterschiedliche Anforderungen einer datengesteuerten Verwaltungstätigkeit erfüllt sowie aussagekräftige und zielgerichtete Ergebnisse schneller zur Verfügung stellen kann.

Tabelle 14: Vor- und Nachteile Tableau Plattform (Quelle: Krohn & Heinze GbR)

Vorteile	Nachteile
Hoher Self Service-Grad	Höherer Schulungsaufwand für jeden Nutzenden durch Self Service
Günstige Katalog-Funktionalität	Kein komplexes Modelling (zentrales Datenmodell)
Kann bereinigte, transformierte Daten in beliebige Ziel-Datenbanken schreiben, welche anschließend als bereinigte, harmonisierte und geprüfte Datenquellen dienen können	Historisierung nur bedingt möglich (manuell)
	Entkopplung der operativen Systeme über Zwischenschicht (zum Beispiel Landing Zone) nötig, aber mit der Komponente „Tableau Prep Builder“ auch in einer Datenbank realisierbar

3.3. Datenanalyse- und Datenvisualisierungstool

Für die Auswahl eines Datenanalyse- sowie eines Datenvisualisierungstools wurden im Rahmen der Interviews und Workshops Anforderungen erhoben, die mit dem Expertenwissen der Syncwork AG ergänzt wurden (vgl. Anhang 2). Aufgrund des Umfangs dieser Anforderungen wurden im Rahmen des Projektzieles „Datenmanagement und Datenanalytik“ elf Kriteriengruppen mit 52 zugehörigen Anforderungen und ebenfalls elf Kriteriengruppen mit 49 Anforderungen für das Projektziel „Datenvisualisierung und BI Dashboard“ definiert. Hierbei kam es bereits zu Überschneidungen bei der Zuordnung von Anforderungen an die Werkzeuge zu den jeweiligen Projektzielen. Dies lässt sich unter anderem mit der Verbindung von Datenanalytik und den (visuellen) Auswertungsmöglichkeiten bei den Ergebnissen begründen. Grundsätzlich bieten alle betrachteten Softwarelösungen eine Möglichkeit der visuellen Auswertung der Ergebnisse. Bei der vorgeschlagenen Auswahl wurde die Empfehlung berücksichtigt, dass die analytischen Funktionen auf die Datenbank ausgelagert werden sollen, um allen Anwendenden und möglichen Anwendungsarten (von der Datenanalyse bis zur Datenvisualisierung) zur Verfügung zu stehen.

Auf der Basis der erhobenen Anforderungen und Erkenntnisse wurde von der Syncwork AG eine Marktbeurteilung vorgenommen und es sind geeignete Softwareprodukte zum Thema Datenanalytik und Datenvisualisierung untersucht worden. Dabei wurden folgende Anwendungen betrachtet:

- Tableau Server
- Oracle Analytics Server
- Microsoft Power BI
- Qlik Sense
- GeoWare InterMonitor
- ESRI ArcGIS Insights
- StatSoft STATISTICA
- Instant Atlas
- Opendatasoft

Insgesamt wurden elf Oberkriterien mit den zugehörigen 52 Unterkriterien zu den Themen Datenanalytik und deren (visuelle) Auswertungsmöglichkeiten erzeugt. Bereits mit dem Informationsauftrag oder dem Erkenntnisziel erfolgt eine Fixierung von Metadaten. Zu nennen sind hier zum Beispiel Rechtsgrundlagen, Definitionen, Erhebungsart/Datenquelle, der Zeit- und Raumbezug und der Merkmalsträger sowie das Sachgebiet, das aus systematischen Gründen eine Rolle spielt. Diese Metadatenobjekte gehören zu der Objektgruppe Definitionsmerkmale, die grundsätzliche Fragen ‚Wer‘, ‚Was‘, ‚Wann‘, ‚Wo‘ und ‚Wie‘ erhoben oder gemessen werden soll, beantwortet.

3.3.1. Datenanalytik

Es wurden Anwendungen untersucht, welche einem überschaubaren Nutzerkreis ermöglichen, die Daten aus den DWH zu analysieren, auszuwerten und zu visualisieren. Hierbei wurden Schwerpunkte definiert und Lösungsempfehlungen erarbeitet. Eine der Herausforderungen ist die Beantwortung der Frage, wie und wo die statistischen Auswertungen erfolgen sollen.

Die Empfehlung, statistische Berechnungen durch teilweise kostenfreie Tools wie in einer Datenbank durchzuführen und somit für alle Nachnutzenden verfügbar zu machen sowie Hemmschwellen bei der aktiven Nutzung von Expertentools gar nicht erst entstehen zu lassen, spiegelt sich auch in der betrachteten Produktauswahl wider. So wurde mit STATISTICA von StatSoft ein reines Statistiktool in die Nutzwertanalyse aufgenommen, während mit Esri, Tableau und Oracle drei eher klassische BI-Tools untersucht und vorgeschlagen wurden.

Alle vier Softwarelösungen bieten eine ausreichende Möglichkeit, die vorhandenen Daten zu analysieren und grafisch ansprechend darzustellen, wobei STATISTICA seine Stärken eindeutig bei der Nutzung analytischer Funktionen hat, die anderen drei Instrumente hingegen bei einer intuitiven Bedienung für die jeweiligen Nutzenden und bei einer größeren Auswahl an Darstellungsmöglichkeiten. Bei der vorgeschlagenen Tool-Aus-

wahl wurde die Empfehlung berücksichtigt, dass die analytischen Funktionen auf die Datenbank ausgelagert werden sollen, um allen Anwendenden und Einsatzmöglichkeiten zur Verfügung zu stehen.

Tabelle 15: Überblick Softwareauswahl für Datenanalytik (Quelle: SyncWork AG)

Softwareauswahl für Datenanalytik	
StatSoft	Die Datenanalytisesoftware STATISTICA von dem Anbieter StatSoft hebt sich durch die umfangreichen Analysemöglichkeiten (univariat, multivariat) von den anderen betrachteten Produkten ab. STATISTICA bietet zusätzliche Datenverwaltungsfunktionen, die für halbstrukturierte oder unstrukturierte Daten (Text, Ton, Videos et cetera) nützlich sind.
Esri	ArcGIS Insights von Esri bietet den Vorteil, große Datenmengen übersichtlich in Form von Karten, Diagrammen und Tabellen darzustellen. Zusätzlich können Daten auf Karten aggregiert werden, um Beziehungen aufzuzeigen und/oder neue Beziehungen aufzudecken. Des Weiteren können die einzelnen Schritte der Analyse in einem Modell dargestellt werden. Sogenannte interaktive Daten-Stories tragen ebenfalls zur Nachvollziehbarkeit bei. Esri ist bekannt für eine intuitive Benutzeroberfläche (per Drag und Drop) und die umfangreiche Produktpalette beziehungsweise das umfangreiche Portfolio der eigenen Anwendungen. Die Einbindung sämtlicher Daten – auch über das DWH – hinaus ist gewährleistet. Neben der Lizenz für ArcGIS Insights ist auch eine Publisher- oder Administratorenrolle in einer ArcGIS-Organisation (ArcGIS-Lizenz) erforderlich. Das setzt den Einsatz von ARCGIS Desktop oder ArcGIS online voraus. Neben räumlichen und nicht räumlichen Untersuchungen können zeitliche Betrachtungen sowie Beziehung- und Regressionsanalysen durchgeführt werden. Zusätzlich ist eine Funktion zur Skripterstellung (Python und R) enthalten.
Tableau	Bei Tableau überzeugt vor allem der Fokus auf eine benutzerfreundliche Oberfläche und die neuesten Visualisierungswerkzeuge. Somit besteht nicht die Gefahr, dass die Nutzbarkeit des Tools durch Komplexität und daraus resultierenden starken Schulungsbedarf eingeschränkt wird, auch wenn grundsätzlich Trainingsmöglichkeiten bestehen.
Oracle	Der Oracle Analytics Server bietet ein umfassend administrierbares Metadaten-Repository. Damit wird ein gezielt vorbereiteter Zugriff auf die Daten ermöglicht, ohne dass sich Endnutzende mit den Strukturen des DWH, den Tabellen und den Spalten auseinandersetzen müssen, um auf die gewünschten Daten zugreifen zu können. Mit dem Analytics-Server ist es zudem nicht nur möglich, auf Daten aus dem DWH zuzugreifen, sondern auch lokal gespeicherte Daten in Analysen zu überführen.

3.3.2. Datenvisualisierung

Durch die Erkenntnisse, dass nicht nur reine Statistik, sondern auch spezialisierte Visualisierungs-Tools (BI-Anwendungen) zu einem gewünschten Ergebnis führen können, erweitert sich die Betrachtungsweise für mögliche Anwendungen enorm. So können neben Visualisierungswerkzeugen, welche Daten für eine größere Anzahl von Nutzenden innerhalb der eigenen Verwaltung zur Verfügung stellen, auch solche betrachtet werden, die bereits einen Schritt weitergehen, nämlich in Richtung eines Open Data-Portals.

Die untersuchten Anwendungen stehen den Visualisierungsmöglichkeiten der Tools nicht nach, jedoch liegt die Stärke dieser Werkzeuge in der Darstellung und Bereitstellung von Daten, welche für diese Softwarelösungen speziell vorbereitet wurden, und nicht in der Erforschung neuer Sachverhalte. Mit Opendatasoft und InstantAtlas lagen zwei dieser möglichen Open Data-Portal Anwendungen auch auf den vorderen Plätzen der gewichteten Nutzwertanalyse.

Tabelle 16: Überblick Softwareauswahl für Datenvisualisierung (Quelle: SyncWork AG)

Softwareauswahl für Datenvisualisierung	
Opendatasoft	Opendatasoft zählt zu den wohl umfangreichsten von den untersuchten Produkten und bietet die größte Vielfalt an Einstellungen, Visualisierungen und Exportmöglichkeiten. Außerdem können optisch anspruchsvolle Dashboards gestaltet werden. Allerdings kann diese Software nicht selbst gehostet werden, sondern läuft ausschließlich auf den Servern des Anbieters, die sich aber auf Wunsch in Europa befinden und somit europäischem Recht unterliegen. Die Daten können über Schnittstellen direkt von Quellsystemen importiert oder über Austauschformate wie CSV-Dateien in Opendatasoft eingelesen werden. Dieses Tool eignet sich hervorragend für den Aufbau eines Open Data-Portals und wird bereits von vielen öffentlichen Stellen in Deutschland und aller Welt genutzt.

Softwareauswahl für Datenvisualisierung	
InstantAtlas	Gegenüber Opendatasoft gehört InstantAtlas zu den Instrumenten, die selbst gehostet werden können. Die Datengrundlage ist hier eine JSON-Datei, die über selbst erstellte Datenbankprozesse direkt aus dem DWH generiert werden kann. Auch für ein angestrebtes Open Data-Portal eignet sich diese Software optisch und lizenztechnisch sehr gut und kann auf eine Vielzahl von Referenzprojekten auch innerhalb von Deutschland verweisen. Der Vorteil bei InstantAtlas liegt in dem geringen Arbeits- und Zeitaufwand, insbesondere bei der Einrichtung. Hinzu kommt, dass bei darzustellenden Daten, sobald sie aufbereitet und eingepflegt sind, eine turnusmäßige Erfassung ohne Probleme in die Online-Ansicht überführt werden kann, da die Datenquelle bereits bekannt ist und lediglich eine Aktualisierung erfolgen muss. Des Weiteren bietet InstantAtlas die Möglichkeit, die auszudruckenden Felder aus- beziehungsweise abzuwählen. Damit lassen sich Berichte auf die gewünschten Inhalte reduzieren und können als Teile für weitere Berichte dienen, da zudem nicht zu erkennen ist, dass diese von einer Website stammen.
Esri	Esri Insights setzt den Einsatz von ARCGIS Desktop oder ArcGIS online voraus. Visualisierungen in Esri Insights können in Karten, Diagrammen und Tabellen dargestellt werden. Es werden diverse Kartentypen unterstützt (Karten mit Abschnitten, Karten mit Einzelwerten, Choroplethenkarten, Säulendiagramm-Symbole, Heatmaps, Beziehungskarten, Positionskarten, Kreisdiagramm-Symbole, Karten mit abgestuften Symbolen). Für die Darstellung in Form von Diagrammen stehen umfangreiche Schaubilder zur Verfügung (Balkendiagramm, gestapeltes Balkendiagramm, Säulendiagramm, gestapeltes Säulendiagramm, Ringdiagramm, Histogramm, Scatterplot, Zeitseriendiagramm, Treemap, Blasendiagramm, Liniendiagramm, Sehnendiagramm, Datenuhr, Heat-Diagramm, Box-Plot, Beziehungsdiagramm, Scatterplot-Matrix, KPI-Kachel, Kombinationsdiagramm, Punktdiagramm). Für die Darstellung und die Organisation in Tabellen sind alle gängigen Funktionen (Filtern, Sortieren) verfügbar. Zusätzlich können Berechnungen durchgeführt und neue Felder hinzugefügt werden. Zu den Funktionen zählen Zeichenfolgefunktionen, numerische Funktionen, Datumsfunktionen und logische Funktionen.

3.4. Fazit

Zusammenfassend lassen sich aus Sicht der Stadt Cottbus/Chósebus folgende Schlussfolgerungen ziehen:

Die Fachbereiche frühzeitig in Maßnahmen hinsichtlich der Etablierung von Data Governance, Data Management oder der Auswahl eines Informationsportals einzubeziehen ist wichtig für die Akzeptanz sowie das Verständnis einzelner Schritte im Gesamtvorhaben. Der Paradigmenwechsel (weg von Datensilos hin zu gemeinsamen Standards und Quellen) ist am ehesten erfolgreich, wenn die Umsetzung des Vorhabens mit begleitender Kommunikation (Change-Management-Maßnahmen) durchgeführt wird. Mithilfe von Test- oder Pilotphasen können die Fachbereiche in der täglichen Arbeit den Nutzen von kommerziellen Produkten bewerten und daraus weitere Bedarfe (zum Beispiel Schulungen im Bereich Datenanalytik, Statistik) ableiten. Ein hoher Grad an Self Service-Fähigkeit der eingesetzten Produkte kann die Umsetzung der Data Governance-Aufgaben in den Fachbereichen gewährleisten. Hierzu sind Schulungen für die einzelnen Rollen der Data Governance-Unit wie den Data Steward oder den Data User vorzusehen. Ziel ist, Aufgaben möglichst eigenständig in den Fachabteilungen zu realisieren.

Soll langfristig ein rechtskonformer Betrieb ohne rechtliche Risiken gewährleistet sein, muss im Rahmen der Planung ein Hosting-Anbieter aus der Europäischen Union, idealerweise aus Deutschland, beauftragt oder ein lokaler Vor-Ort-Betrieb genutzt werden. Cloudbasierte Ansätze der US-Firmen bieten skalierbaren Zugriff auf robuste, intelligente und benutzerfreundliche Funktionen und ermöglichen einen sofortigen Start, aber die Rechtsunsicherheit ist für eine deutsche Kommune aktuell ein zu großer Unsicherheitsfaktor. Derzeit gibt es keine funktional gleichwertigen rechtssicheren Cloud-Alternativen innerhalb von Europa oder Deutschland, welche nicht von US-Firmen betrieben werden.

Die Umsetzung eines zentralen agilen DWH nach der Data Vault-Methodik ist bezüglich der allgemeinen und individuellen Anforderungen der Stadt Cottbus/Chósebus anzustreben.

Als wesentliche allgemeine Anforderungen wurden formuliert:

- Ein zentrales, harmonisiertes Datenmodell
- Die vollständige Historisierung der Fachbereichsdaten

- Die Entkopplung der operativen Systeme von Analyse- und Reportsystemen
- Eine durchgängige Nachvollziehbarkeit der Datenherkunft (Data Lineage)
- Die weitestgehende Automatisierung der Datentransformationen

Zu den individuellen Anforderungen an den Datenintegrationsprozess zählen insbesondere:

- Die schrittweise, additive Integration der Daten der einzelnen Fachbereiche
- Die Auflösung der existierenden Datensilos
- Die Etablierung von Data Governance-Aufgaben in den Fachbereichen
- Die zentrale Datenbereitstellung für eine Vielzahl von Anwendungen (nicht ausschließlich BI-Tools)
- Integration und Betrieb der Systeme in der kommunalen IT-Infrastruktur

Die Auswahl des geeigneten Datenbankmanagementsystems (DBMS), der passenden Data Warehouse-Automatisierungslösung (DWA) und des BI-Toolset ist aufeinander abzustimmen und sollte in einem End-to-End Proof of Concept (POC) validiert werden, das heißt von den Datenquellen über die Integrationslösung bis zu den auswertenden Systemen. Eine erste Checkliste für einen solchen POC wurde im Rahmen des Leitfadens „Datenplanung“ erarbeitet.

Der Zugang zu Analyse- und Statistikanwendungen ist einem möglichst breiten Benutzerkreis in den Fachbereichen zu ermöglichen. Daher wird auf einen hohen Self Service-Grad der Analyse-Tools gesetzt. Die untersuchten Self Service BI-Tools ermöglichen sowohl einfache explorative Datenanalysen mit anschaulichen visuellen Darstellungen als auch spezielle analytische Auswertungen wie Zeitreihenanalysen. Ihr Einsatz ermöglicht, die Datenquellen umfassend zu analysieren, Datentransformations- und Bereinigungsverfahren zu erarbeiten sowie einfache fachspezifische dimensionale Datenmodelle für die Auswertung zu erstellen.

Dabei profitieren sie in großem Maße von einem vorgelagerten Data Warehouse, welches als „alleinige“ Datenquelle den Zugang zu bereits integrierten und aufbereiteten Daten gemäß eines gemeinsamen Datenmodells bietet. Mit einem Data Warehouse nach Data Vault ergeben sich die Möglichkeiten, einerseits die Raw-Marts Analysen über den historisierten Rohdaten zu fahren und andererseits über die Information-Marts Zugriff auf spezifische, aufbereitete und verdichtete Datenzusammenstellungen zu erhalten. Die Bereitstellung spezifischer Data Marts im Data Warehouse bietet die optimale Nutzung der Analyse- und Statistikfunktionen des verwendeten DBMS. Die untersuchten DWA-Tools ermöglichen eine modellbasierte Generierung der Data Marts und bieten damit einen hohen Automatisierungsgrad.

Literaturverzeichnis

1. Leitfaden Datenerhebung:

AG, A.-S. (Januar 2010). Konzept zum Aufbau eines Infrastrukturknotens der kreisfreien Stadt Cottbus. Cottbus.

Brandenburg, L. (1. April 2020). Brandenburgisches Statistikgesetz – BbgStatG. Brandenburg.

DCAT-AP.de. (8. Oktober 2020). DCAT-AP.de Spezifikation 1.1. Von dcat-ap.de: <http://dcat-ap.de/def/dcatde/1.1/spec/specification.pdf> abgerufen

Erlhöfer, D. (4. Dezember 2019). Datenvorbereitung für Qlik Sense und Tableau. Von b.telligent: <https://www.btelligent.com/blog/datenvorbereitung-fuer-qlik-sense-und-tableau/> abgerufen

Mikrodaten. (18. September 2018). Von Wikipedia: <https://de.wikipedia.org/wiki/Mikrodaten> abgerufen

Porzelt, A. M. (Dezember 2014). Hinweise zur Anonymisierung von qualitativen Daten.

Tableau. (Februar 2021). Das Tableau-Datenmodell. Von <https://help.tableau.com>: https://help.tableau.com/current/pro/desktop/de-de/datasource_datamodel.htm abgerufen

Zeiler, G. (2017). Von <https://www.sigs-datacom.de/>: http://www.sigs.de/publications/bi/2017/DataWarehouseAutomation/zeiler_BIS_OTS_DWA_2017.pdf abgerufen

2. Leitfaden Datenqualität:

Daniel Linstedt, M. O. (2016). 10.5 Implementing the Error Mart. In M. O. Daniel Linstedt, Building a Scalable Data Warehouse with Data Vault 2.0.

Spilka, S. (3. Oktober 2019). DATENINTEGRITÄT: BEDEUTUNG, RISIKEN & DER RICHTIGE UMGANG. Von <https://anexia.com/blog/de/datenintegritaet/> abgerufen

WANDT, D. H. (14. September 2015). Datenqualität. Von www.digitalwiki.de/datenqualitaet/: <http://www.digitalwiki.de/datenqualitaet/> abgerufen

3. Leitfaden Datenqualität:

„Standardklauseln“ nicht haltbar – Heise Newsticker. (4. August 2020). Von <https://www.heise.de/news/Google-setzt-auf-Standarddatenschutzklauseln-nach-gekipptem-Privacy-Shield-4862466.html> abgerufen

BARC. (2020). BARC Guide, BI & Analytics 2021. Von BARC Business Application Research Center: www.barc.de/guides abgerufen

Beilhammer, M. (Februar 2021). Data Vault: Datenmodellierung für agile Anwendungen. Von industryPRESS: <https://www.industry-press.com/data-vault/> abgerufen

BI-SURVEY.com – Comparison of the Best Business Intelligence and Analytics Software Products in 2021. (Februar 2021). Von <https://bi-survey.com/business-intelligence-software-comparison> abgerufen

CLOUD Act – iX 07/2018. (kein Datum). Von <https://www.heise.de/select/ix/2018/7/1530927567503187> abgerufen

CLOUD Act – iX 07/2018. (2018). Von <https://www.heise.de/select/ix/2018/7/1530927567503187> abgerufen

Datavault BUILDER. (Februar 2021). Von <https://datavault-builder.com/de/features/> abgerufen

Digitale Verwaltung: Open Source und offene Standards fürs E-Government. (19. September 2020). Von Heise Newsticker: <https://www.heise.de/news/Digitale-Verwaltung-Open-Source-und-offene-Standards-fuers-E-Government-4906207.html> abgerufen

Gartner. (17. August 2016). Use the Integration Maturity Model to Assess and Improve Your Integration Competency.

Ist-Analyse_Kriterienkatalog_Daten_FB33.xlsx. (September 2020).

Kippung „Privacy Shield“ – Heise Newsticker. (16. Juli 2020). Von <https://www.heise.de/news/Was-das-Ende-des-Privacy-Shield-fuer-den-Datenverkehr-mit-den-USA-bedeutet-4852929.html> abgerufen

Piatkowski, D. (März 2015). Data Vault – Agilität und Skalierbarkeit im Enterprise Data Warehouse. Von Oraylis.de: <https://www.oraylis.de/blog/data-vault-agilitaet-und-skalierbarkeit-im-enterprise-data-warehouse> abgerufen

Leitfaden Datenvisualisierung

B. Saket, A. Endert and Ç. Demiralp, „Task-Based Effectiveness of Basic Visualizations,“ in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 7, pp. 2505-2512, 1 July 2019, doi: 10.1109/TVCG.2018.2829750.

Quelle: Statistisches Bundesamt (Destatis), 2021 | Stand: 07.05.2021 / 08:02:22 - (Code: 21111-0010)

Statistisches Bundesamt Deutschland – GENESIS-Online: Die Datenbank des Statistischen Bundesamtes ([destatis.de](https://www.destatis.de))

Abbildungsverzeichnis

Abbildung 1: DWH-Schichten nach Data Vault (Krohn & Heinze GbR)	13
Abbildung 2: Zentrales Data Warehouse als Integrationsschicht	15
Abbildung 3: Aspekte der Datenqualität je Datenhaltungsebene (Krohn & Heinze GbR)	17
Abbildung 4: Operative Metadaten nach Data Vault (Krohn & Heinze GbR)	17

Tabellenverzeichnis

Tabelle 1: Überblick über die verschiedenen Kompetenzprofile	7
Tabelle 2: Gewichteter Anforderungskatalog der Fachbereiche der Stadt Cottbus zur städtischen Data Governance	10
Tabelle 3: Vor- und Nachteile des Top-Down-Ansatzes bei der Erstellung des zentralen Datenmodells für das Core Data Warehouse	11
Tabelle 4: Vor- und Nachteile des Bottom-Up-Ansatzes bei der Erstellung des zentralen Datenmodells für das Core Data Warehouse	12
Tabelle 5: Hauptziele der Datenerhebung	14
Tabelle 6: Herausforderungen des Datenmanagements	15
Tabelle 7: Anforderungen an das Datenqualitätsmanagement	16
Tabelle 8: Kriterien der Datenauswertung	18
Tabelle 9: Überblick verschiedener Diagrammartentypen	20
Tabelle 10: Übersicht über die visuell-retinalen Variablen nach ihrer Eignung	21
Tabelle 11: Vor- und Nachteile Datavault Builder™	35
Tabelle 12: Vor- und Nachteile Qlik Compose	36
Tabelle 13: Vor- und Nachteile Qlik Sense	36
Tabelle 14: Vor- und Nachteile Tableau Plattform	36
Tabelle 15: Überblick Softwareauswahl für Datenanalytik	38
Tabelle 16: Überblick Softwareauswahl für Datenvisualisierung	38