

Datenmanagement



Arbeitsgruppe

Marc Godat

Dipl. Experte in Rechnungslegung und Controlling
Lostorf

Marek Gossner

Dipl. Experte in Rechnungslegung und Controlling
Wallisellen

Peter Herger

Inhaber des eidg. Fachausweises im Finanz- und Rechnungswesen
Bäretswil

Daniel Keller

MAS Accounting & Finance
Rafz

Dr. Abbas Tutcuoglu

PhD Space Engineering, California Institute of Technology
Wallisellen

Markus Zäch

Dipl. Maschineningenieur HTL
Eggersriet

Manuel Zmak

Dipl. Experte in Rechnungslegung und Controlling
Bülach

Inhaltsverzeichnis

Vorwort	5
1 Einleitung	6
1.1 Definition	6
2 Ziele und Grundlagen des Datenmanagements	9
2.1 Einführung Datenbanken	9
2.2 Zentrale Begriffe	9
3 Über die Form der Daten	13
3.1 Datenformate	13
3.2 Struktur der Daten	13
3.3 Granularität der Daten	15
3.4 Bereinigung von heruntergeladenen Daten	15
4 MS Access – ein Überblick	16
4.1 Tabellen	16
4.2 Beziehungen	16
4.3 Abfragen	17
4.4 Formulare, Berichte, Seiten und Module	17
5 Aufbau eines Data Warehouse mit MS Access	18
6 ERP-Systeme (Enterprise-Resource-Planning-Systeme)	20
6.1 Was ist ein ERP-System?	20
6.2 Datenintegration durch ERP-Systeme.....	21
6.3 Historische Entwicklung von ERP-Systemen	22
6.4 Auswahl eines ERP-Systems	23
6.5 ERP-Software-Anbieter und -Lösungen	24
6.6 ERP-Hersteller	24
6.7 Kosten eines ERP-Systems	24
6.8 Aufbau des ERP-Systems/ERD	25
6.9 Grenzen und Perspektiven moderner ERP-Systeme	28
7 Business Intelligence (BI)	30
7.1 Definition von Business Intelligence	30
7.2 Warum ist Business Intelligence wichtig?	31
7.3 Warum Business Intelligence nutzen?	31
7.4 Business-Intelligence-Software	32
7.5 Online Analytical Processing (OLAP)/Business Analytics	32
7.6 Reporting und Abfragen	33
7.7 Digitale oder visuelle Dashboards	33
7.8 Integration	33

8	Big Data	34
8.1	Wo wird Big Data genutzt?.....	34
8.2	Big Data auswerten	34
8.3	Big Data nutzen	35
8.4	Wo speichert man grosse Datenmengen?	35
8.5	Big-Data-Softwarelösungen	35
8.6	Kritik an Big Data	36
9	SQL	37
9.1	Einführung	37
9.2	Befehl SELECT	37
9.3	Befehl DISTINCT	38
9.4	Befehl WHERE	38
9.5	SQL-Operatoren AND/OR	39
9.6	SQL IN Operator	40
9.7	SQL-Operator BETWEEN	41
9.8	Befehl SQL LIKE	42
9.9	SQL ORDER BY Operator.....	43
9.10	Befehl SQL GROUP BY	43
9.11	Befehl SQL HAVING	44
10	Robotic Process Automation (RPA)	46
10.1	Einleitung	46
10.2	Aufbau	46
10.3	Robotic Process Automation bei einer Kundendatenbank	47
10.4	Robotic Process Automation bei Softwaretests.....	48
10.5	Einrichtung von Robotic Process Automation – was gilt es zu beachten?	48
11	Künstliche Intelligenz und Business Analytics	50
11.1	Was ist künstliche Intelligenz?	50
11.2	Die verschiedenen Formen der KI	52
11.3	Klassifizierung und Regression	53
11.4	Die Entstehung von KI-Modellen	54
11.5	Anfälligkeiten eines KI-Modells	61
11.6	Der schnelle Weg zum ersten Modell	62
12	Digitale Transformation und neue Rollenbilder	64
12.1	Digitalisierung	64
12.2	Zielsetzungen einer digitalen Transformation	65
12.3	Herausforderungen der digitalen Transformation.....	67
12.4	Roadmap einer Digital Finance Transformation	67
13	Veränderung der finanziellen Führung und Steuerung der Organisation	70
13.1	Neue Rollenbilder in der digitalen Welt der Finanzfunktionen	70
13.2	Change-Management	73
14	Literaturverzeichnis	76
15	Abbildungsverzeichnis	77

Vorwort

Sehr geehrte Damen und Herren

Seit Jahrzehnten gibt es die hochspezialisierten, eidgenössischen Berufsprüfungen «Fachausweis im Finanz- und Rechnungswesen» und «dipl. Experte/Expertin in Rechnungswesen und Controlling». Diese gehören zu den ältesten eidgenössischen Berufsprüfungen, welche sich über die Jahre halten konnten. Die Grundlage für das Rechnungswesen ist auch noch heute die im Jahre 1494 vom italienischen Mathematiker Luca Pacioli erstmals beschriebene doppelte Buchhaltung.

In den 1970er startete die Digitalisierung des Rechnungswesens, da die Buchhaltung als logische Abfolge relativ einfach in elektronischer Form erfasst und dargestellt werden konnte. Auch heute ist es darum noch üblich, dass die gesamte Informationstechnologie dem CFO angehängt oder unterstellt ist und er die Verantwortung für die dazugehörigen digitalen Prozesse trägt.

Daten sind das neue Öl, und die Prozesse, welche die Daten generieren, die neuen Raffinerien. Umso wichtiger ist es, die Datenhoheit klar zu bestimmen und die Verantwortung für Daten und Prozesse zentral zu regeln. Aus diesem Grunde wurden die eidgenössischen Berufsprüfungen im Rechnungswesen mit dem neuen Fach Datenmanagement ergänzt. In diesem Fach wird die Grundlage für den richtigen Umgang mit den Daten und das Verständnis für Prozesse erarbeitet, damit künftige Absolventinnen und Absolventen für die Zukunft gerüstet sind.

Leider fehlt es nach wie vor an Fachliteratur, so dass diese Broschüre als Hilfsmittel für den zukünftigen «dipl. Experte/Expertin in Rechnungswesen und Controlling» verfasst worden ist.

Herzlichen Dank an die Autoren und an den Verein für die höheren Prüfungen in Rechnungswesen und Controlling, ohne deren Engagement die Realisierung dieser Broschüre nicht möglich gewesen wäre. Die Zusammenarbeit in der Fachkommission Datenmanagement, mit grosser Unterstützung des Vereins für die höheren Prüfungen in Rechnungswesen und Controlling, war sehr fruchtbar und ich bin stolz, in welcher kurzen Zeit wir eine so breite und informative Broschüre realisieren konnten.

Den zukünftigen Absolventinnen und Absolventen unserer Berufsprüfungen wünsche ich viel Erfolg. Ich bin überzeugt, mit dem Rucksack «Datenmanagement» wird das Rechnungswesen in der Schweiz auch in Zukunft modern bleiben, so dass sich dieses in dieser schnell lebigen Zeit kontinuierlich weiterentwickeln kann.

Zürich, im November 2021

Herzlich

Peter Herger

Vorstandsmitglied veb.ch, Verantwortlicher Digitalisierung Rechnungswesen und Controlling.

1 Einleitung

Der Begriff «Datenmanagement» ist vielseitig interpretierbar. Der vorliegende Standard entspricht dem Verständnis der Fachkommission Datenmanagement, die aus Praktikern aus dem Bereich Rechnungswesen und Controlling besteht. Die nachfolgenden Darlegungen beanspruchen nicht, das weitreichende Thema Datenmanagement vollumfänglich zu erfassen, sondern es soll anhand eines Beispiels – einer Kundendatenbank – einen Faden durch die verschiedenen Thematiken des Datenmanagements ziehen.

1.1 Definition

Folgende Begriffe werden in diesem Standard mit der angegebenen Bedeutung in alphabetischer Reihenfolge verwendet. Die in den jeweiligen Absätzen erklärten Begriffe werden hier nicht redundant definiert.

Begriff	Beschreibung
APICS	The Association for Operations Management (APICS) ist eine 1957 in den USA gegründete Organisation, die sehr einflussreich Belange der Betriebsorganisation gestaltet, ausbildet und auch zertifiziert.
APS	Advanced Planning and Scheduling-Systeme (APS-Systeme) werden seit Beginn der 2000er-Jahre zur Erweiterung der Planungsfunktionen nach der MRP-II-Methode eingesetzt und dienen der Optimierung von Lieferterminen. Eine allgemein anerkannte und exakte Definition besteht zurzeit nicht, weshalb jeder ERP-Hersteller eigene Modelle entwickelt.
Data Warehouse (DWH)	Eine für Auswertung aufgebaute Datenbank. Oftmals wird das DWH aus verschiedenen Datenquellen gefüttert.
DTA/SEPA	Für den Inlandszahlungsverkehr besteht ein Datenträgeraustausch-Format (DTA oder DTAUS-Format). Dieser einheitliche Standard ermöglicht die elektronische Verarbeitung von Zahlungsaufträgen (Überweisungen und Lastschriften) im schweizerischen Inlandszahlungsverkehr. DTA wird sukzessive durch den SEPA-Standard (Single Euro Payments Area [SEPA]) abgelöst.
EDI	EDI (Electronic Data Interchange) bewerkstelligt den elektronischen Datenaustausch von Geschäftsdokumenten zwischen Unternehmen. Die normierte EDI-Definition ist ein elektronisches Standardformat, das Papierdokumente wie Bestellungen oder Rechnungen ersetzt.
ELM	Die zertifizierte Software «Einheitliches Lohnmeldeverfahren» (ELM) regelt und organisiert die Übermittlung von Lohndaten zwischen Unternehmen, Sozialversicherungen und Behörden. Dabei werden Lohndaten an AHV, Versicherer, Steuerämter und das Bundesamt für Statistik übermittelt.
ERD	Die grafische Darstellung von Beziehungstypen (Entitäten) wird Entity-Relationship-Diagramm (ERD) genannt. Die Darstellung ist eine grafische Übersicht der internen Zusammenhänge einer Datenbank oder eines Datensystems, wodurch ein komplex erscheinendes, netzartiges Gebilde entstehen kann.
ERP	ERP steht für Enterprise Resource Planning und definiert die Inhalte einer Softwarelösung zur Ressourcenplanung eines Unternehmens bzw. einer Organisation. ERP-Systeme integrieren eine Vielzahl von Geschäftsprozessen, die in einer zentralen Datenbank verarbeitet und gespeichert werden.

Begriff	Beschreibung
ESR	In der Schweiz werden als Einzahlungsschein mit Referenznummer (ESR) zwei verschiedene Formulare verwendet, um einen Geldbetrag in Schweizer Franken zu überweisen. Der ESR wird ab 2020 durch die QR-Rechnung abgelöst.
ICT	ICT steht in den Bereichen Elektronik und Computertechnik für Information and Communications Technology, also für Informations- und Kommunikationstechnik. Hinweis: ICT hat diverse weitere Bedeutungen in anderen Fachbereichen.
ISO 20022	ISO 20022 ist der internationale Standard auf Basis von XML-Meldungen (Extensible Markup Language) für den elektronischen Datenaustausch in der globalen Finanzbranche. Der Schweizer Finanzplatz hat diesen Standard eingeführt und so den Zahlungsverkehr vereinfacht und damit einen höheren Automatisierungsgrad der Zahlungsprozesse erreicht. Das herkömmliche DTA-File wurde in diesem Zusammenhang per 30.6.2018 durch die pain.001-Meldung nach dem ISO 20022-Standard abgelöst.
MRP II	Manufacturing Resources Planning (MRP II) ergänzt die Produktionsprogrammplanung mit weiteren Planungsebenen (Absatzplan), welche die rollende Planung und Budgetierung erleichtern. Das Modell der unbeschränkten Kapazitäten wurde durch Kapazitätsgrenzen nach dem Modell der Rückwärtsterminierung optimiert.
MRP und MRP I	Material Requirements Planning (MRP) beschreibt die Materialbedarfsplanung mehrstufiger Stücklisten und die Berechnung von Brutto- und Nettobedarf. Die Ende der 1970er-Jahre entwickelte MRP I berücksichtigt neben den Stücklisten auch die Termine von Arbeitsgängen. Das damit verbundene Modell unbeschränkter Kapazitäten hat sich in der Praxis nie richtig durchgesetzt, weshalb MRP II entwickelt wurde.
PPS	PPS ist ein computergestütztes Produktionsplanungs- und -steuerungssystem. Das PPS-System ist zentraler Bestandteil eines ERP-Systems, weil damit die Planung finanzieller und personeller Ressourcen des Betriebs ausgeführt werden kann.
Produktionsprogramm	Das Produktionsprogramm enthält die Gesamtheit aller Produkte eines Unternehmens in der Art und Menge und bestimmt zugleich den Ort der zu produzierenden Produkte für einen festgelegten Zeitpunkt bzw. definierten Zeitraum.
Ressource	Eine Ressource (Quelle) beschreibt Mittel, Gegebenheit wie auch Merkmal bzw. Eigenschaft, um Ziele zu verfolgen, Anforderungen zu bewältigen, spezifische Handlungen zu tätigen oder einen Vorgang zielgerecht ablaufen zu lassen. In der Betriebswirtschaft, Volkswirtschaft und in Organisationen werden darunter Betriebsmittel, Finanzen, Rohstoffe, Energie sowie Personen und (Arbeits-)Zeit verstanden.
RFID	RFID (Radio Frequency Identification), die Identifizierung mithilfe elektromagnetischer Wellen, bezeichnet eine Technologie für Sender-Empfänger-Systeme zum automatischen und berührungslosen Identifizieren und Lokalisieren von Objekten und Lebewesen mit Radiowellen.

Begriff	Beschreibung
SCOR	Supply Chain Operations Reference Model (SCOR) beschreibt alle unternehmensinternen und -externen Geschäftsprozesse. Seit 2014 ist SCOR in APICS integriert. Die vorliegenden Unterlagen berücksichtigen die Grundlagen nach dem SCOR-Modell.
Simultaneous Engineering	Methode zur Verkürzung und Minimierung von Innovationszyklen.
Software Deployment	Neue oder geänderte Software durchläuft einen Deployment-Prozess, d. h., die Software wird installiert, konfiguriert und in der Produktionsumgebung ausgerollt.
SSEE	SSEE ist das Gegenteil von VUCA, nämlich stable (beständig), secure (sicher), easy (einfach), explicit (klar) und beschreibt die Voraussetzungen für einfachere Unternehmensführung.
State of the Art	Entspricht der aktuellen Technologie, die Umsetzung erfolgte nach Best Practices.
Supply Chain	Die Supply Chain (Wertschöpfungskette, logistische Kette, Lieferkette, Versorgungskette) eines Produkts umfasst alle wertgenerierenden Aktivitäten sämtlicher Fertigungs- und Absatzstufen, angefangen bei den Rohmaterialien bis hin zu den fertigen Produkten.
VUCA	VUCA ist ein Akronym für die englischen Begriffe volatility (Unbeständigkeit), uncertainty (Unsicherheit), complexity (Komplexität) und ambiguity (Mehrdeutigkeit). Es beschreibt schwierige Rahmenbedingungen der Unternehmensführung in Zeiten der Globalisierung.

2 Ziele und Grundlagen des Datenmanagements

2.1 Einführung Datenbanken

In der heutigen Zeit ist Excel im Buchhaltungs- und Controlling-Bereich omnipräsent. Man findet in den entsprechenden Abteilungen teilweise richtige «Excel-Monster», die eine enorme Menge an Daten verarbeiten. Sehr oft wird Excel hier für Tätigkeiten eingesetzt, für die es nicht entwickelt wurde. Man muss sich immer bewusst sein, dass Excel keine Datenbank ist und auch nicht als solche genutzt werden sollte (auch wenn die immer leistungsfähigeren Computer dies durchaus möglich machen). Was ist hierzu die Empfehlung von Microsoft, wann resp. bis wohin Excel und ab wann Access genutzt werden soll?

"Use **Access** when you:

- Require a **relational** database (multiple tables) to store your data.
- May need to **add more tables** in the future to an originally flat or non-relational data set
- Have a **very large amount of data** (thousands of entries)
- Have data that is mostly of the **long text string type** (not numbers or defined as numbers).
- Rely on **multiple external databases** to derive and analyze the data you need
- Want to run **complex queries**
- Have **many people working** in the database and want robust options to expose that data for updating"

"Use **Excel** when you:

- Require a flat or **non-relational** view of your data (you do not need a relational database with multiple tables)
- Want to run primarily **calculations** and **statistical comparisons** on your data
- Know your **dataset is manageable in size** (no more than 15,000 rows)."

Zusammenfassend:

MS Excel eignet sich eher für Aufgaben mit begrenzter Komplexität, mit Zweidimensionalität, mit eher geringen Datenmengen, mit hohen Ansprüchen an Flexibilität und geringen Integritätsanforderungen.

MS Access eignet sich eher für die Verarbeitung von grossen Datenmengen, insbesondere dann, wenn diese mehrdimensional und relationaler Natur sind (siehe zentrale Begriffe unten).

2.2 Zentrale Begriffe

Im Folgenden werden einige zentrale Begriffe aus der «Datenbank-Welt» erläutert. Es wirkt etwas theoretisch, allerdings wird einiges sehr schnell Schritt für Schritt klarer werden, wenn Sie im Skript weiter vorankommen.

Cubetools

Unter Cubetools verstehen wir eine Softwareapplikations-Kette (Software-Suite), die aus einer Access-Datenbank, einem oder mehreren OLAP-Cubes und einem oder mehreren Frontends besteht. Die Software-Suite arbeitet mit Daten, die aus vorgelagerten Systemen übernommen werden. Die Architektur kann grafisch folgendermassen dargestellt werden:

OLAP-Cube

Ein OLAP-Cube (auch Data-Cube, Cube-Operator oder OLAP-Würfel genannt) ist ein in der Data-Warehouse-Theorie gebräuchlicher Begriff zur logischen Darstellung von Daten. Die Daten werden dabei als Elemente eines **mehrdimensionalen** Würfels (engl. cube) angeordnet. Die Dimensionen des Cube beschreiben die Daten und erlauben auf einfache Weise den Zugriff. Daten können über eine oder mehrere Achsen des Würfels ausgewählt werden. Die Bezeichnung OLAP stammt aus der Datenanalyse und bedeutet Online Analytical Processing.

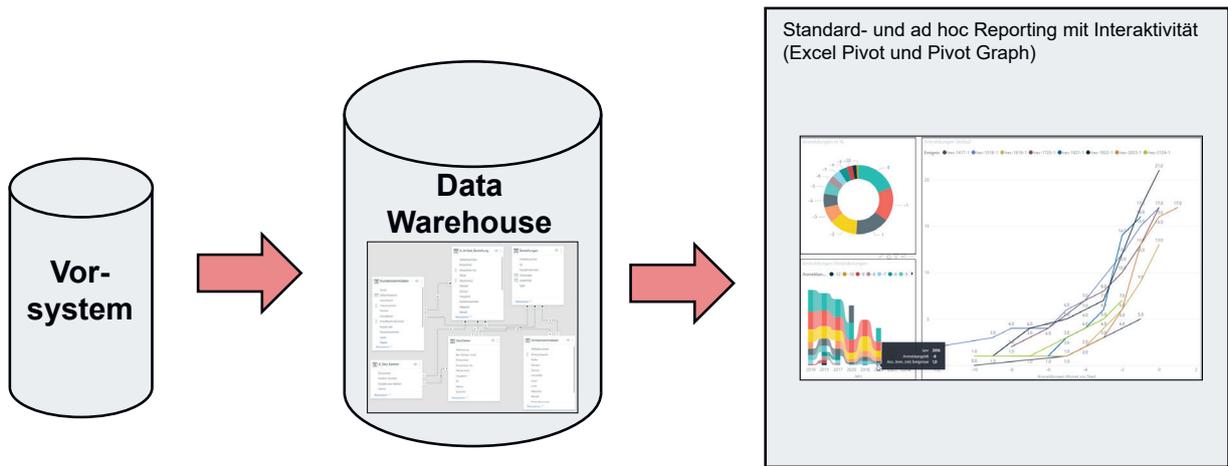


Abbildung 1: Die Architektur von Cubetools

Diese Art der Darstellung ist für die Analyse von Daten von Vorteil, da auf verschiedene Aspekte (Dimensionen) der Daten auf gleiche Weise zugegriffen wird. Daher auch der Einsatz bei OLAP-Anwendungen, die die Daten in einem Data Warehouse analysieren oder visuell aufbereiten. (Quelle: Wikipedia)

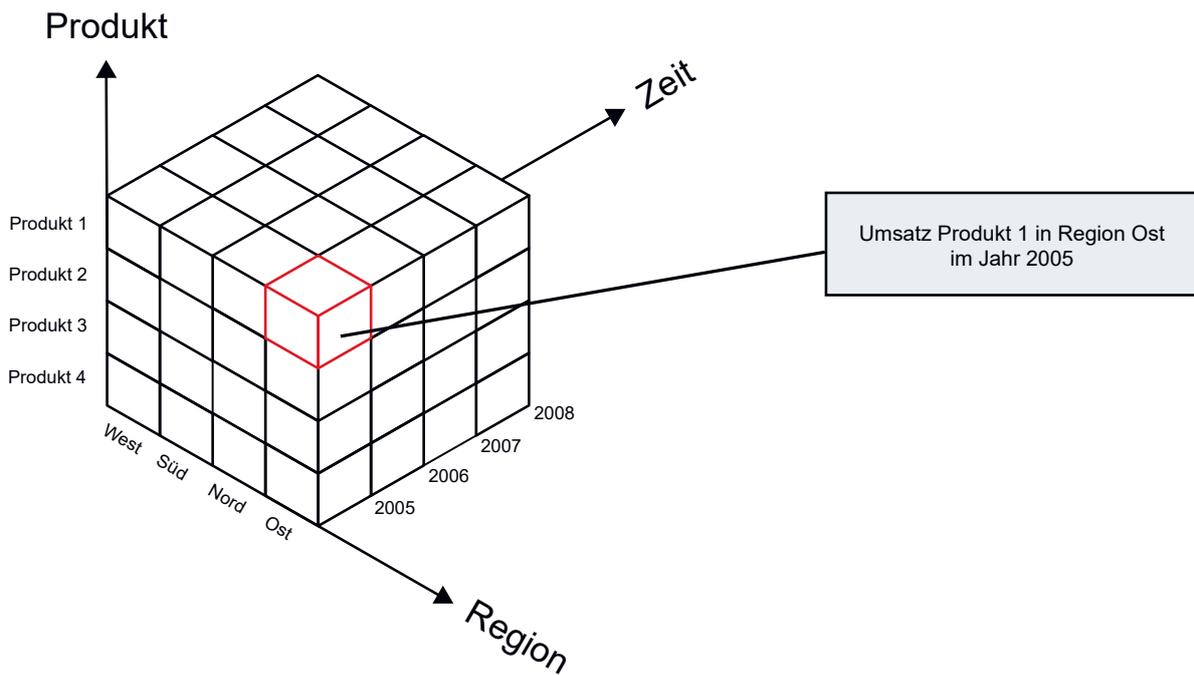


Abbildung 2: Beispiel OLAP-Cube

Data Warehouse (DW)

Ein Data Warehouse (dt. Datenlager) ist eine zentrale Datensammlung (meist eine Datenbank), deren Inhalt sich aus Daten unterschiedlicher Quellen zusammensetzt. So kann ein Controlling Data Warehouse etwa aus einem Teil Umsatzstatistik, einem Teil Finanzbuchhaltung, einem Teil Lagerbewirtschaftung, einem Teil Forecast – die alle aus einem ERP-System kommen können – und einem Teil Offertwesen – der aus einem CRM-System kommen kann – bestehen. Die Daten werden von den Datenquellen in das Data Warehouse geladen und dort vor allem für die Datenanalyse und zur betriebswirtschaftlichen Entscheidungshilfe in Unternehmen langfristig gespeichert. Der Begriff stammt aus dem Informationsmanagement in der Betriebswirtschaft. Ein Data Warehouse dient der Informationsintegration.

Es gibt derzeit keine einheitliche Definition für den Begriff «Data Warehouse». Weitgehend einig ist man sich in Folgendem:

Ein Data Warehouse ermöglicht eine globale Sicht auf heterogene und verteilte Datenbestände, indem die für die globale Sicht relevanten Daten aus den Datenquellen zu einem gemeinsamen konsistenten Datenbestand zusammengeführt werden. Somit entsteht der Inhalt eines Data Warehouse durch Kopieren und Aufbereiten von Daten aus unterschiedlichen Quellen. **Meist ist ein Data Warehouse die Basis für die Aggregation von betrieblichen Kennzahlen und Analysen innerhalb mehrdimensionaler Matrizen** (OLAP-Cube), dem sog. Online Analytical Processing (OLAP). Ein Data Warehouse ist häufig Ausgangsbasis für Data Mining (auch Datenschrüfung). In der Regel arbeiten die Anwendungen mit anwendungsspezifisch erstellten Auszügen aus dem Data Warehouse, den sog. Data Marts. (Quelle: Wikipedia)

Normalisierung

Bei der Normalisierung werden Spalten von Tabellen innerhalb der Datenbank in neue Spalten aufgeteilt (z. B. Adressen in Postleitzahl, Ort und Strasse) oder diese Spalten mit anderen Tabellen verknüpft (z. B. ein Kunde über eine eindeutige Nummer mit einer Kundentabelle).

Der **Sinn der Normalisierung** besteht darin, Redundanzen (gleiche, mehrmals vorhandene Informationen) zu verringern und Anomalien (z. B. einander widersprechende Dateninhalte) zu verhindern, um so die Wartung einer Datenbank zu vereinfachen sowie die Konsistenz der Daten zu gewährleisten.

Deutlich wird dies an einem einfachen **Beispiel**: Eine Datenbank enthält Kunden und deren Adressen sowie Aufträge, die den Kunden zugeordnet sind. Würde jeder Auftrag direkt die Adresse des Kunden speichern, müsste man bei einer Adressänderung alle Datensätze nach diesem Kunden durchsuchen und die Adresseinträge aktualisieren. Bei einer normalisierten Datenbank gibt es dafür nur einen einzigen Eintrag, hier die Kundentabelle. Zudem ist der Speicherbedarf geringer, wenn der Datensatz einer Tabelle (z. B. «Auftrag») auf einen Datensatz einer anderen Tabelle (z. B. «Kunde») verweist, anstatt diese Daten selbst zu enthalten.

Da es mehrere Aufträge vom gleichen Kunden geben könnte, würden ansonsten je Auftrag immer wieder die Daten des Kunden kopiert werden (Redundanz). Gäbe es ausserdem in der Datenbank mehrere Adressen zu einem Kunden, könnte es zu widersprüchlichen oder falschen Ausgaben kommen – die Daten wären dann nicht konsistent. (Quelle: Wikipedia)

Integrität

Datenintegrität beschreibt die Anforderung an den Schutz und die Qualität von digitalen Daten. Um die Datenintegrität sicherzustellen, muss die Konsistenz, Vollständigkeit, Genauigkeit und Gültigkeit von Daten über den gesamten Aufbewahrungszeitraum garantiert sein. Alle Datenänderungen müssen nachvollziehbar dokumentiert werden, sodass Daten nicht unbemerkt oder unberechtigt verändert werden können. Datenintegrität hat damit das übergeordnete Ziel, Daten vor internen und externen Verstößen zu schützen.

Stammdaten

- zeichnen sich durch eine gewisse Statik aus (zeitlich invariant) und haben meistens keinen Zeitbezug,
- werden oft von mehreren Anwendungen bzw. Unternehmensbereichen verwendet, z. B. Teilstammsätze (Einkauf, Konstruktion, Disposition, Buchhaltung, Vertrieb, Arbeitsvorbereitung),
- sind bei analytischen Auswertungen oft die Kriterien, nach denen ausgewertet wird (z. B. Produkt, Filiale, Kunde), womit sie Kandidaten für die Dimensionen im Online Analytical Processing (OLAP) sind, und
- werden meistens langfristig gehalten.

Bewegungsdaten

- zeichnen sich durch eine gewisse Dynamik aus (zeitlich variant) und haben meistens einen Zeitbezug (Gültigkeitsdatum),

- werden oft von wenigen Anwendungen genutzt,
- liefern oft die Fakten bei analytischen Auswertungen, womit sie Kandidaten für die Zelleninhalte im OLAP sind,
- werden meistens nur zeitlich begrenzt benötigt und daher
- kurz oder mittelfristig gehalten.

3 Über die Form der Daten

Ausgangslage ist, dass ein (oder mehrere) Systeme bestehen, die Daten beinhalten, mit denen der Verantwortliche arbeiten möchte. Die Vielfalt solcher Systeme (sie werden hier als **Vorsysteme** bezeichnet) und der Strukturen ihrer Daten ist sehr gross; somit müssen allgemeine Hinweise zu den wichtigsten Punkten ausreichen.

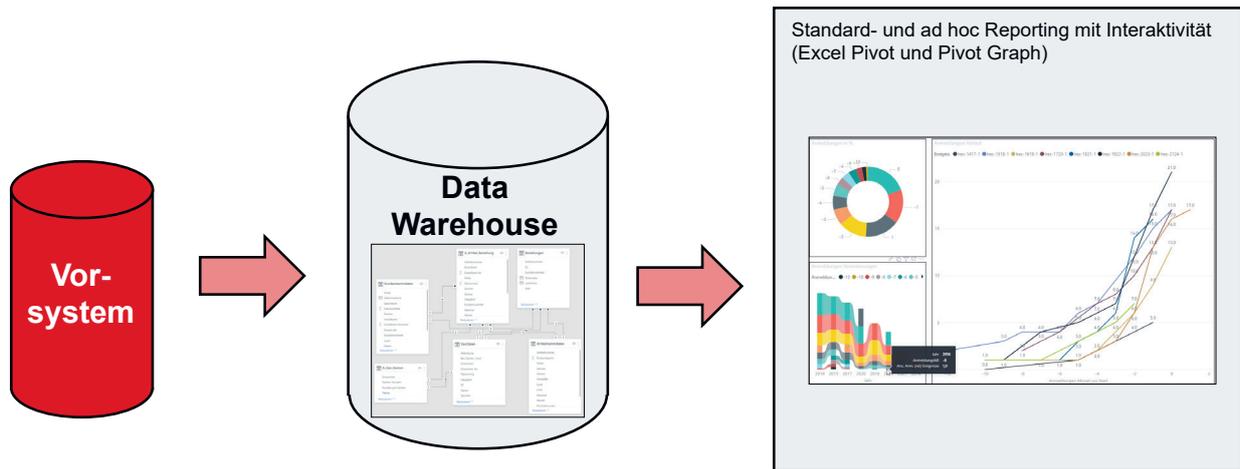


Abbildung 3: Vorsystem und Data Warehouse

3.1 Datenformate

MS Access kann mit folgenden Datenformaten arbeiten: MS Access, dBASE III/IV/V, MS Excel, Exchange, HTML, Lotus 1-2-3, Outlook, Paradox, Textdateien, XML, ODBC-Datenbanken. Aus mehr oder weniger allen bestehenden Datensystemen (wie z. B. SAP) können Daten in einem oder mehreren von diesen Formaten heruntergeladen werden. Manchmal ist eine Zwischenstufe erforderlich; so kann es durchaus Sinn machen, SAP-Daten zunächst als Textfile aus SAP heraus zu exportieren, sie dann in Excel zu **bereinigen** und das Excel-File anschliessend in MS Access zu importieren.

3.2 Struktur der Daten

Grundsätzlich müssen die Daten aus den Vorsystemen in **normalisierter** Form vorliegen. Für ein Data Warehouse braucht man Datensätze mit einem Feld für jedes relevante **Merkmal**. Alle Merkmale, mit denen man arbeiten möchten, müssen Bestandteil eines Datensatzes sein. Eine strukturierte Erfolgsrechnung und Bilanz aus einem Buchhaltungssystem kann beispielsweise folgende Form haben (Detail):

Erfolgsrechnung Februar 2020 Year-to-date				
Buchungskreis (Firma)	Kostenartengruppe	Kostenartennummer und -name	Wert aktuelles Jahr	Wert Vorjahr
BU21	Personalkosten			
		450400 Saläre	47'123.–	45'231.–
		450300 Pensionskasse	13'652.–	13'211.–
		450450 Halbtax-Abo	0.–	450.–
	Personalkosten gesamt		60'775.–	58'892.–

Die abgebildete Form ist als Grundlage für die Übernahme von Daten in ein Data Warehouse aus folgenden Gründen **nicht geeignet**:

- Der **Buchungskreis** (BU21) ist in der Regel für Analysen wichtig, ist hier jedoch nicht Bestandteil eines Datensatzes.
- Der **Kostenartenname** ist nicht notwendig, da dieser in einer 1:1-Beziehung zu der Kostenartennummer steht und somit über eine Stammdatentabelle Kostenarten herbeigezogen werden kann (**Normalisierung**).
- Eine Aufteilung der Werte in **Aktuell und Vorjahr** ist nicht zweckmässig, weil es sich vom Wesen des Feldinhaltes her um das Gleiche handelt, und zwar den Wert der dieser Erfolgsrechnung zugrunde liegenden Transaktionen. Zweckmässiger erscheint eine Zuordnung des Feldes zu einem **Buchungsdatum**.
- Eventuell möchte man bei einer Auswertung auch die Möglichkeit haben, **Drilldowns** bis auf Stufe einzelne Buchung zu machen. Dies würde bedeuten, dass man hier statt zusammengefasster Werte die einzelnen **Transaktionen** herbeiziehen müsste.
- Möchte man zudem wissen, welche Personalkosten im Januar und welche im Februar angefallen sind, sowohl im aktuellen Jahr wie auch im Vergleichsjahr, würde dieses Anliegen ebenfalls durch das **Buchungsdatum** gelöst werden.
- Eine **Summe** der Personalkosten braucht man nicht, zumal diese als Summe aller Datensätze (oder aller Datensätze mit einem oder mehreren gemeinsamen Merkmalen) in einem Data Warehouse gebildet werden.

Die **Lösung** könnte aus einem Datensatz mit folgenden Feldern bestehen:

Feldname	Beispiel Feldinhalt	Kommentar
Buchungskreis	BU 21	Über eine Stammdatentabelle im Data Warehouse kann der Buchungskreis mit einem Firmennamen verbunden werden, was Auswertungen leserlicher macht.
Kostenart	450400	Hier genügt die Kostenartennummer, weil der Name der Kostenart über eine entsprechende Stammdatentabelle herbeigeht werden kann.
Buchungsdatum	14.02.2012	Mit dem Buchungsdatum als Merkmal kann logisch abgeleitet werden, in welchem Monat, Quartal und Jahr eine Transaktion erfolgswirksam war.
Buchungstext	Rückstellung Bonus anteilig	Wenn man die Möglichkeiten hat, Drilldowns bis auf Stufe Einzeltransaktion zu machen, liefert der Buchungstext nützliche Zusatzinformationen für Analysezwecke.
Wert	3'000.–	Wert der einzelnen Transaktion. Dieser kann in der Datenbank über alle Dimensionen der Datentabelle aggregiert werden.

Falls man selbst nicht das Wissen hat, um eigene, massgeschneiderte Abfragen in einem VORSYSTEM zu erstellen, können diese bei der Softwarelieferantin in Auftrag gegeben werden. Die Kosten dafür halten sich in der Regel in Grenzen, da auch die VORSYSTEME nach den Grundsätzen der Datenbankentwicklung mit normalisierten Datenbeständen aufgebaut werden. So sind zusammenfassende Auswertungen in aller Regel mit normalisierten Daten als Ausgangspunkt erstellt worden.

3.3 Granularität der Daten

Die Wahl der Granularität der Daten ist eine Abwägung zwischen **Auswertungsmöglichkeiten** einerseits und **Performance/Geschwindigkeit** andererseits. MS Access kann ein ansehnliches Datenvolumen verarbeiten; wenn allerdings mehrere Hunderttausend bis zu einer Million Datensätze zusammenkommen, sollte man nach Optimierungsmöglichkeiten Ausschau halten. So kann es am Beispiel der Finanzbuchhaltung durchaus sinnvoll sein, im aktuellen Jahr auf Stufe der einzelnen Transaktionen zu arbeiten; wenn die Daten dann im Laufe der Zeit zu Vorjahreswerten werden, braucht es vielleicht nur noch die Saldi pro Monat. In einem solchen Fall können die Einzeltransaktionen **kondensiert** werden, das heisst, man summiert sie nach sinnvollen Kriterien und reduziert somit die Anzahl Datensätze.

3.4 Bereinigung von heruntergeladenen Daten

Auch dann, wenn in den Vorsystemen die notwendigen Abfragen für das Abrufen von normalisierten Datensätzen vorhanden sind, fallen nach der Datenextraktion häufig kleinere Bereinigungsarbeiten an. Mögliche Gründe dafür könnten sein, dass die Vorsystemabfrage mehr Felder liefert, als man für das Data Warehouse braucht, oder die Formate der Feldinhalte angepasst werden müssen.

So liefert beispielsweise **SAP** standardmässig Daten, bei denen Zahlenwerte über 1000 mit einem Komma getrennt werden. Kommawerte kann MS Access jedoch nicht lesen; solche Daten verursachen beim Datenimport in das Data Warehouse Fehlermeldungen.

Da Datenimporte in ein Data Warehouse regelmässig (monatlich, wöchentlich oder auf Abruf) gemacht werden müssen, ist eine manuelle Bereinigung der Daten aufwendig und mit einem gewissen Risiko für **Fehlmanipulationen** behaftet. Es empfiehlt sich, eine solche Datenbereinigung möglichst mittels **Excel-Makros** vorzunehmen.

4 MS Access – ein Überblick

Eine MS Access-Datenbank besteht aus unterschiedlichen Objektarten, die im Folgenden kurz erläutert werden.

4.1 Tabellen

Eine Tabelle ist **der Ort, wo die Daten physisch gespeichert werden**. Sie besteht aus **Feldern**. Die **Art der Felder** wird nach dem vorgesehenen **Inhalt** des Feldes festgelegt. Eine Zeile mit zwei oder mehreren Feldern ist ein **Datensatz**. Die gebräuchlichsten Feldarten sind:

Feldart	Kommentar
Text	Dieser Feldtyp wird in der Praxis bei allen Feldern angegeben, die nicht zu den übrigen drei Typen gehören. Die maximale Feldlänge beträgt 255.
Zahl	Dieser Feldtyp wird bei allen Feldern eingesetzt, die in Zahlenberechnungen eingesetzt werden sollen.
Datum/Uhrzeit	Dieser Feldtyp wird für Datumsangaben eingesetzt. Mit diesem Feldtyp können Werte über Zeitperioden aggregiert werden.
AutoWert	Der AutoWert ist ein Zähler, der innerhalb einer Tabelle die Einmaligkeit eines Datensatzes gewährleistet. Jedem neu hinzugefügten Datensatz wird eine Laufnummer erteilt. Falls der Datensatz gelöscht wird, erlischt mit ihr auch die Laufnummer und wird nicht für einen später hinzugefügten Datensatz verwendet. Aus diesem Grund wird der AutoWert in der Regel als Primärschlüssel eingesetzt.

Der **Primärschlüssel** ist dasjenige Feld in der Tabelle, welches sicherstellt, dass der Datensatz einmalig ist. In einer Kundendatenbank könnte dies etwa die AHV-Nummer sein, da diese für jeden Menschen einmalig ist. Verfügt man nicht über einen einmaligen Wert, kann man auch einen Zähler (AutoWert) einsetzen.

Einen Primärschlüssel braucht es dann nicht, wenn die Einmaligkeit eines Datensatzes bereits durch ein Vorsystem gewährleistet ist. So ist es in einem Buchhaltungssystem z.B. nicht möglich, zwei Kostenarten mit der gleichen Kostenartennummer, aber mit unterschiedlichen Bezeichnungen zu hinterlegen. In einem solchen Fall muss auch in der Stammdatentabelle Kostenarten im Data Warehouse kein Primärschlüssel definiert werden.

4.2 Beziehungen

Die Beziehungen definieren das **Verhältnis** der einzelnen Tabellen in einer Datenbank zueinander. Beim Thema Normalisierung wurde das Beispiel einer Datenbank mit Kunden- und Auftragsdaten erwähnt. Es wurde erklärt, warum es sinnvoll ist, die Kundendaten in einer Tabelle und die Auftragsdaten in einer anderen Tabelle zu speichern.

Weil ein einziger Kunde mehrere Aufträge erteilen kann, umgekehrt jedoch ein Auftrag immer nur von einem Kunden stammen kann, hat man eine **1:n-Beziehung** zwischen der Kundentabelle und der Auftrags-tabelle. Dies ist die gebräuchlichste Beziehungsart in einer Datenbank.

Eine Datenbank mit definierten Beziehungen zwischen Tabellen wird als **relationale Datenbank** bezeichnet.

4.3 Abfragen

Eine Abfrage ist eine **Darstellung** von Daten, die in einer Tabelle gespeichert sind.¹ Mittels Abfragen können Daten selektiert, gruppiert, kombiniert und in eine Form gebracht werden, die für Analysezwecke optimiert ist.

Eine Abfrage kann entweder

- direkt auf einer Tabelle oder
- auf einer anderen Abfrage basieren.

Im letzteren Fall sprechen wir von einer **Abfragehierarchie**.

4.4 Formulare, Berichte, Seiten und Module

Formulare werden dazu eingesetzt, um die Dateneingabe und Datensuche direkt in MS Access zu vereinfachen. Mit der **Berichtsfunktion** können Daten in vordefinierten Berichtsformaten dargestellt werden. Die **Seiten** haben die gleiche Funktion wie die Formulare, nur in einem Website-Format. Mit den **Modulen** können eigene Programme bzw. Programmteile geschrieben werden.

¹ Dies trifft für die Abfragearten *Auswahlabfrage* und *Kreuztabellen-Abfrage* zu.

5 Aufbau eines Data Warehouse mit MS Access

Den umfassendsten Teil der Entwicklungsarbeit stellt sicher der Aufbau eines Data Warehouse innerhalb eines Cubetool-Entwicklungsprojekts dar.

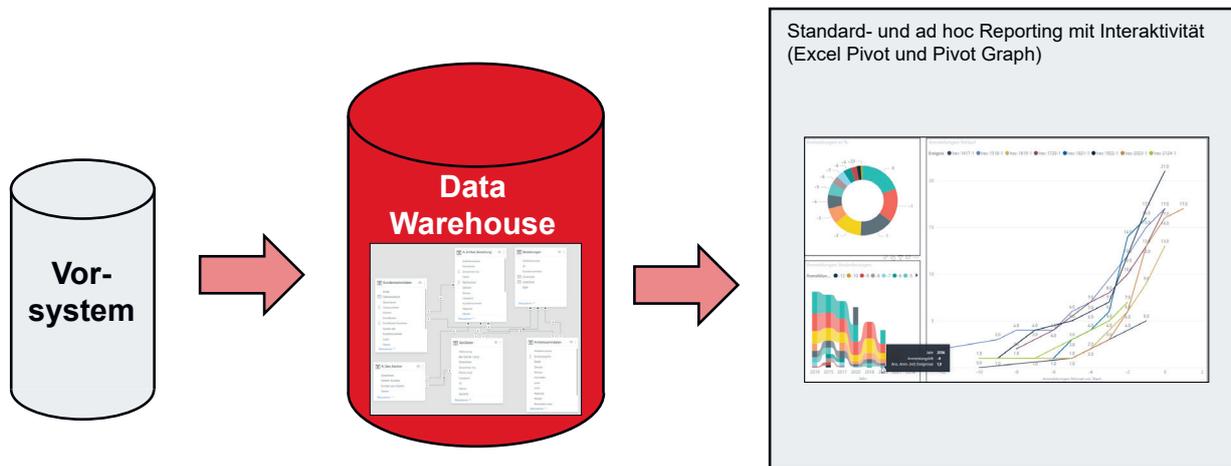


Abbildung 4: Cubetool – Data Warehouse

Das Vorbild für das hier gezeigte Beispiel ist die angepasste Version einer Datenbank, die in drei Geschäftseinheiten (ca. 400 MA, ca. CHF 100 Mio. Umsatz) einer bedeutenden Firmengruppe seit mehreren Jahren erfolgreich eingesetzt wird. Das Controlling der drei Geschäftseinheiten ist zentralisiert. Als Data Warehouse dient Access, in welchem die Daten aus den Vorsystemen (SAP) verarbeitet und anschliessend via Excel Frontend ausgegeben werden.

Dieses Cubetool hat nach Aussage des Leiters Controlling den Aufwand für die Aufbereitung von Daten für Controllingzwecke **um rund 90 % reduziert**. Die betreffende Datenbank wurde als Data Warehouse konzipiert und beinhaltet ausschliesslich Tabellen und Abfragen – keine Formulare, Berichte oder Makros. Als Frontend dient MS Excel (Pivot Graph).

Ins Data Warehouse werden nach jedem Monatsabschluss Ist-Werte aus einem **SAP-Vorsystem** in Form einer **Kundenauftragspositionsrechnung** (Erträge und direkte Kosten) und einer **Profitcenterrechnung** (vollständige Erfolgsrechnung) mit Einzelpositionen übernommen. Des Weiteren werden bei jeder Planungsrunde – Forecast 1, 2 und 3 (IST + Vorschau auf das Jahresende) sowie Target Setting (3 Jahre) und Budget (3 Jahre) – Planungsaktualitäten angelegt. Diese Planungsaktualitäten (inkl. Budget) werden nicht in SAP eingespeist, sondern direkt im Data Warehouse gespeichert. In das Data Warehouse werden nur Erfolgsrechnungsdaten und keine Bilanzdaten eingelesen.

Im Folgenden wird die Struktur der Tabellen und Abfragen des Data Warehouse aufgezeigt. Die drei unteren Datenbankobjekte **T_Planungsaktualitaeten_neu**, **T_KE5Z_mit_Kundauftr** sowie **T_KE24_neu** sind Tabellen, die übrigen Datenbankobjekte sind Abfragen. Die Pfeile zeigen den Fluss der Daten in der Abfragehierarchie. In den Kästchen sind die Namen der Tabellen und Abfragen hinterlegt.

In Abbildung 5 sieht man den systematischen Aufbau eines solchen Data Warehouse. Im linken Sektor werden die Plandaten eingespeist, im mittleren und rechten Sektor die Ist-Daten. Im oberen Bereich werden die Ist-Daten mit den Plandaten zusammengeführt und durch die Stammdaten angereichert. Die oberste Abfrage bildet schliesslich die Basis für ein Excel Frontend.

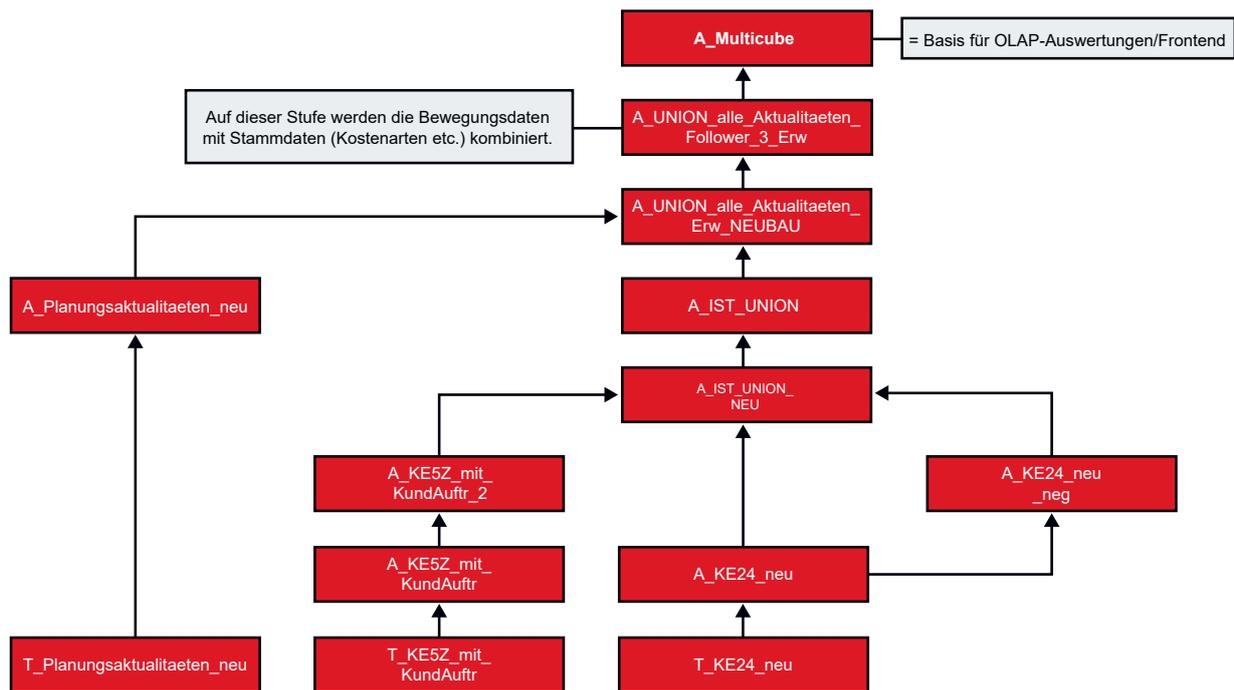


Abbildung 5: Beispiel systematischer Aufbau eines Data Warehouse

Der grosse Vorteil einer solchen Architektur ist sicher im einfachen Unterhalt begründet. Die Datenanreicherung resp. Aktualisierung erfolgt nur in den drei Tabellen **T_Planungsaktualitaeten_neu**, **T_KE5Z_mit_Kundauftr** sowie **T_KE24_neu**. An den restlichen Abfragen muss man – vorausgesetzt, dass keine strukturellen Änderungen stattfinden – nichts mehr ändern. Diese verarbeiten die hochgeladenen Daten automatisch.

An diesem Beispiel sieht man recht anschaulich, wie ein KMU auf eine sehr sparsame Weise (nur mit vorhandenen Mitteln) ein Management Informationssystem (MIS) auf Basis eines Data Warehouse erstellen kann.

6 ERP-Systeme (Enterprise-Resource-Planning-Systeme)

Unternehmen sind in den letzten Jahrzehnten immer mehr gewachsen. Es sind riesengrosse Konzerne entstanden mit sehr unübersichtlichen Strukturen, enorm grossen Cashflows und dadurch auch einem sehr grossen Kapitalbedarf. Solche Konzerne sind schon in normalen Zeiten sehr komplex zu führen. Doch in der heutigen Zeit, in der sich das Umfeld immer schneller ändert, die Lebenszyklen von Produkten und Trends immer kürzer werden, ist es mit gängigen Werkzeugen nahezu unmöglich, solche Firmengebilde zu überblicken. Hier kann man eine Brücke zum ERP-System schlagen. Durch ihre Fähigkeit, das Unternehmen zeitnah in seiner Gesamtheit abzubilden, sind Enterprise-Resource-Planning-Systeme aus der heutigen Entscheidungspraxis kaum mehr wegzudenken.

6.1 Was ist ein ERP-System?

Die meisten Leute, vor allem im Finanzwesen, haben schon von ERP-Systemen gehört. Doch was macht ein ERP-System eigentlich? Grundsätzlich bildet ein ERP-System die im Unternehmen vorhandenen Ressourcen ab, um dem Management die richtigen, zeitnahen Informationen zu liefern und dieses zu steuern. Zu den Ressourcen zählen beispielsweise:

- Kapital
- Mitarbeiter
- Betriebsmittel

Das Hauptziel eines ERP-Systems ist die Verbesserung der organisatorischen Abläufe und Strukturen, schnellere Anpassungsfähigkeit an Unternehmens- und Marktveränderungen sowie die Optimierung der Geschäftsprozesse.

ERP-Systeme sind integrierte Programme, die auf einer zentralen Datenverwaltung basieren. Dadurch werden die Informationsflüsse innerhalb eines Geschäftsprozesses unterstützt, abgebildet, optimiert und dokumentiert.

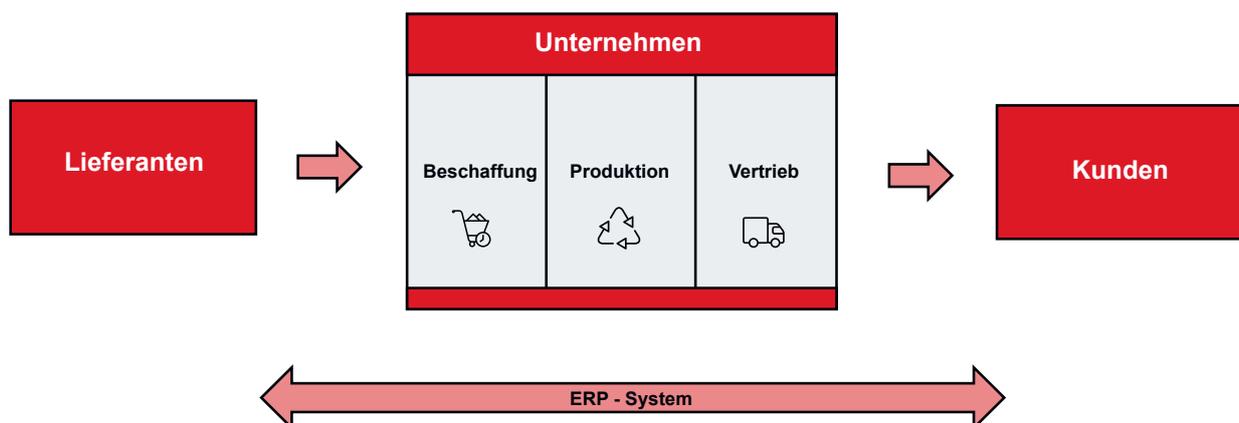


Abbildung 6: Prozess-Sicht eines ERP-Systems

Die grosse Neuerung gegenüber älteren Systemen ist beim ERP, dass diese oben genannten Prozesse in ihrer ganzen Kette abgebildet werden. Dies ermöglicht dem Management, das Unternehmen in seiner Gesamtheit zu erfassen und zu analysieren.

Dies ermöglicht dem Management, wesentlich schneller auf Marktveränderungen zu reagieren und somit mögliche negative Auswirkungen auf das Geschäftsergebnis abzufangen resp. abzufedern.

Organisatorisch betrachtet, setzt sich ein ERP-System aus nachfolgenden Unternehmensbereichen zusammen. Durch das integrative Design verfügen alle Bereiche über dieselbe Datenbasis und kommunizieren untereinander.

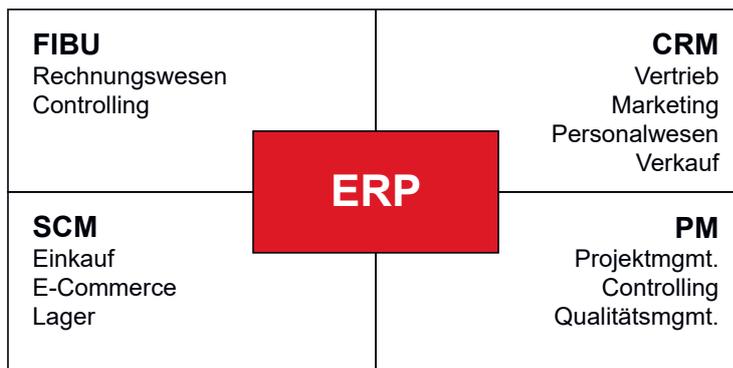


Abbildung 7: ERP-Organisation

FIBU: **Finanzbuchhaltung**

CRM: **Customer-Relationship-Management.** Durch eine systematische Ausrichtung der internen Unternehmensprozesse und mithilfe spezieller Software (CRM-Systeme) sollen die Akquisition von Neukunden und die Pflege von Bestandskunden vorangetrieben und optimiert werden. Ziel ist es, Kundenbeziehungen systematisch aufzubauen und langfristig zu erhalten.

WaWi: **Bezeichnet ein softwaregestütztes Warenwirtschaftssystem,** das die Warenströme eines Unternehmens über die verschiedenen Handelsstufen abbildet. Zentrale Bestandteile sind der Verkauf, Einkauf, Wareneingang und Warenausgang sowie die Lagerhaltung. Warenwirtschaftssysteme können Angebote erstellen, mittels Workflow-Komponenten Aufträge und Bestellungen vollautomatisch abwickeln sowie Lieferscheine und Rechnungen bearbeiten. Mit einer entsprechenden Komponente zur Lagerverwaltung können Warenbestände proaktiv gesteuert und Inventuren durchgeführt werden. Oft sind Warenwirtschaftssysteme in ERP-Systemen als integraler Bestandteil vorhanden (teilweise als zusätzliches Modul aufpreispflichtig) und können vergangene Warenlieferungen, aktuelle Lagerbestände und zukünftig benötigte Waren für Bestellungen erfassen und diese automatisch verarbeiten.

PPS: **Produktionsplanungs- und Steuerungssystem.** Bei einem PPS-System erfolgt die Planung für die Produktion softwaregestützt. Dadurch ist es Unternehmen möglich, mit dieser Form der Planung Kosten im Bereich der Produktion zu senken und die Lieferbereitschaft zu erhöhen. Durch eine PPS-gestützte Optimierung können Unternehmen die Durchlaufzeit ihrer Produktion reduzieren, eine gleichmässige Auslastung der Maschinen erreichen, die Lieferbereitschaft und Termintreue erhöhen und die Lagerbestände und somit auch die Kapitalbindung reduzieren und optimieren.

6.2 Datenintegration durch ERP-Systeme

Die zeitnahe Speicherung von unvorstellbar grossen Datenmengen, ihre Erfassung und Analyse stellt eine grosse Herausforderung für moderne ERP-Systeme in Grossunternehmen und Konzernen dar. Sämtliche Prozesse, die im Hintergrund ablaufen und eine reibungslose Datenverarbeitung sicherstellen, sind einerseits enorm komplex und andererseits enorm kostenintensiv. Ganze Abteilungen sind mit der Aufgabe betraut, eine reibungslose Datenverarbeitung zu garantieren.

Obwohl ein ERP-System aus zahlreichen Softwaremodulen besteht, kann es quer über alle Branchen seine Aufgabe nur dann erfüllen, wenn die Daten unternehmensweit eindeutig und nur einmal vorhanden sind. Die Informationen, die aus diesen Daten gewonnen werden, stellen für ein Unternehmen das höchste Gut dar. Darum sind integrale Daten wichtig und können andernfalls grossen Schaden und Reibungsverluste bewirken.

6.3 Historische Entwicklung von ERP-Systemen

Einzelnen Funktionsbereiche von Unternehmen wie Vertrieb, Material- oder Personalwirtschaft wurden lange Zeit getrennt voneinander betrieben und durch sogenannte Insellösungen unterstützt. Diese Insellösungen lieferten, isoliert betrachtet, zwar absolut korrekte Daten, jedoch wurden durch die fehlende Vernetzung oft inkonsistente Entscheidungen getroffen, die zu Verlusten (monetär oder zeitlich) führten.

Die ersten ERP-ähnlichen Systeme (damals hatten sie noch einen anderen Namen) fand man in den 1960er-Jahren in der Industrie. Diese Systeme wurden hauptsächlich eingesetzt, um Fertigungsprozesse zu optimieren. Die sogenannten MRP-Systeme (Material Resource Planning), die als die ersten Vorläufer der heute gebräuchlichen ERP-Systeme gelten, wurden dazu genutzt, um eine standardisierte Materialbedarfsplanung zu erreichen.

Die Einführung dieser Systeme brachte grosse Effizienzgewinne und man fragte sich bald, ob man diese Systeme nicht ausbauen kann, um auch andere Unternehmensbereiche miteinzubeziehen. Dieses Vorhaben wurde in den 1980er-Jahren erreicht, als die MRP-Systeme um die Produktionsplanung erweitert wurden – man spricht hier von sogenannten MRP-II-Systemen. Es kamen dann mit der Zeit immer mehr Funktionen wie etwa das Personalwesen, CRM und Logistik hinzu, und bald sprach man erstmals von ERP-Systemen.

Die ERP-Systeme wuchsen kontinuierlich, konnten immer mehr Daten verarbeiten und immer mehr Bereiche abdecken. Die heutigen modernen ERP-Systeme werden oftmals auch als ERP II-Systeme bezeichnet, die sogenannte zweite Generation. Sie umfassen zusätzlich zu den internen Prozessen auch unternehmensübergreifende Bereiche entlang der Wertschöpfungskette und basieren auf einer Webarchitektur. Funktionen wie Customer Relationship Management (CRM) und Supply Chain Management (SCM) sind in die Funktionen des ERP-Systems integriert. Somit wird das Bearbeiten von Prozessen möglich, ganz gleich, ob sie intern oder extern stattfinden. Heute ist es beispielsweise mit einem modernen ERP-System problemlos möglich, extern initiierte Aufträge zu übernehmen und diese unternehmensintern zu verarbeiten.

Im Weiteren sind durch innovative Lösungen der ERP-Hersteller Möglichkeiten entstanden, ERP-Systeme auch KMU zugänglich zu machen. Hier sind die neu entwickelten Cloud-ERP zu erwähnen. Bei der ERP-Lösung aus der Cloud (Software-as-a-Service, SaaS) bieten die Anbieter ausgereifte und oft mehrstufige Lösungen für mittelständische Unternehmen an. Neben der ausgereiften, standardisierten ERP-Lösung genießt das Unternehmen zum Beispiel auch State-of-the-Art-Lösungen im Bereich Datensicherheit, die das in den Unternehmen selbst vorgehaltene Sicherheitslevel oft um ein Vielfaches übersteigen.

ERP aus der Cloud ist ein noch relativ junges Thema und hat insbesondere hinsichtlich Datenverfügbarkeit und Datensicherheit mit Vorurteilen zu kämpfen. Für kleine Unternehmen, die sich weder die Software noch die Hardware, geschweige denn eine IT-Abteilung leisten können, kann es jedoch eine effiziente und relativ kostengünstige Lösung sein, um systemtechnisch mit den Grossunternehmen mithalten zu können.

Neben dem Trend SaaS- und Cloud-ERP beherrscht Mobile Computing das Thema ERP. Das BYOD-Phänomen (Bring Your Own Device) kommt dem Bedürfnis vieler Mitarbeiter und Unternehmen entgegen, mit ihren privat genutzten Smart-Gadgets – sei es ein Smartphone, ein Tablet-PC oder ein Notebook – auch betrieblich arbeiten zu können und von überallher auf die Unternehmensdaten zugreifen zu können. Mit dem Mobile-ERP auf Cloud-Basis werden ERP-Systeme noch flexibler und lassen sich noch passgenauer an die Bedürfnisse der Unternehmen und Mitarbeiter anpassen.

In jüngerer Zeit hat zusätzlich eine Entwicklung Einzug gehalten, bei der ERP-Hersteller auch Social-Media-Elemente in ihre Systeme integrieren. Durch diese Integration wandelt sich das ERP-System zu einem Werkzeug, das auch die lebendige Interaktionsbereitschaft und -fähigkeit der Mitarbeiter als Geschäftsprozess und Ressource begreift. Gerade die jüngeren Generationen wurden in eine von Technik bestimmte Welt hineingeboren und nutzen die sozialen Plattformen mit grosser Selbstverständlichkeit. Oft stellen sie die bevorzugte Art und Weise dar, wie Informationen generiert, bereitgestellt, ausgetauscht und weitergeleitet werden.

Immer mehr ERP-Anbieter greifen diesen Trend auf und integrieren Social-Media-Elemente in ihre Softwarelösungen. Informationen können so zentral bereitgestellt und überall zu jeder Zeit dort verteilt werden, wo und wann sie benötigt werden. Personenbezogene Informationen aus den sozialen Medien können im CRM integriert und gesammelt werden, um sie beispielsweise als Grundlage für die Erstellung neuer Kampagnen heranzuziehen.

In letzter Zeit ist zunehmend die Tendenz aufgetaucht, dass es für ERP-Hersteller durch die steigende Komplexität immer schwieriger wird, allen Ansprüchen von Firmen bezüglich CRM, WaWi und PPS gerecht zu werden. Dieser Umstand öffnete die Tür für spezialisierte Softwarehersteller, die sich hier einen Nischenplatz sichern konnten, indem sie sich sehr geschickt auf ein einzelnes Modul spezialisierten und so in der Lage sind, State-of-the-Art-Lösungen anzubieten. Als prominentes Beispiel ist hier sicher das Softwareunternehmen Salesforce zu nennen, das sich auf voll anbindbare CRM-Lösungen und Cloud Computing spezialisiert hat und diese Systeme sehr erfolgreich im Markt anbietet.

6.4 Auswahl eines ERP-Systems

Angesichts von etwa 250 unterschiedlichen ERP-Standardlösungen und geschätzt rund 1000 branchenspezifischen Lösungen für die betriebswirtschaftliche Software allein in der Schweiz, stellt die Auswahl der passenden ERP-Lösung oft eine grosse Hürde für ein Unternehmen dar.

Fällt die Entscheidung zugunsten der Neueinführung eines ERP-Systems, empfiehlt es sich, auf einen standardisierten Evaluationsprozess zurückzugreifen – und falls man unternehmensintern nicht über das nötige Know-how verfügt, sich dieses extern zu beschaffen. Über Erfolg oder Misserfolg entscheiden in dieser Phase meist die richtigen Fragen, gepaart mit ausreichend Zeit, um diese sorgfältig zu beantworten. In diesen Prozess müssen neben der Geschäftsleitung auch andere Verantwortliche und Know-how-Träger miteinbezogen werden.

Welche Kriterien für ein Unternehmen im Einzelnen zum Tragen kommen, hängt unter anderem davon ab, ob man sich für eine On-Premise-Lösung, das heisst Festinstallation auf dem unternehmenseigenen Server, oder für eine Software-as-a-Service-Lösung in der Cloud entscheidet. Nicht jedes Unternehmen möchte seine Daten auslagern und auch die Bindung an einen Anbieter wird oft skeptisch betrachtet, vor allem im Hinblick auf Auswahl, Betrieb und Wartung einer ERP-Lösung. Viele befürchten, sich durch den Erwerb sowie Lizenz- und Folgekosten finanziell zu übernehmen.

Wie lange ist die Nutzungsdauer von ERP-Software?

ERP-Wechselzyklen zeigen an, wie lange Unternehmen an ihren ERP-Systemen festhalten, bevor sie auf neue Lösungen umsteigen. Hierbei sind nicht Release-Wechsel gemeint, sondern Umstiege auf ein komplett neues System. Tendenziell werden die ERP-Wechselzyklen immer kürzer.

Ein wesentlicher Grund für diesen Trend, also die immer kürzer werdende Bindungsdauer des bisherigen ERP-Systems, ist in einem Paradigmenwechsel zu suchen. Dies liegt daran, dass sich immer mehr Unternehmen für Cloud-ERP-Lösungen entscheiden. Im Gegensatz zu On-Premise-Lösungen (Festinstallation auf eigenen Servern), die zunächst aufwendig installiert werden müssen, ist die Cloud-ERP-Software sofort einsatzbereit.

Weitere Faktoren für die immer kürzer werdenden Wechselzyklen sind zudem die erleichterte Datenübernahme, die kürzeren Einarbeitungszeiten und ein Trend zu fallenden Kosten.

6.5 ERP-Software-Anbieter und -Lösungen

Da ERP-Systeme heutzutage längst nicht mehr nur auf Grossunternehmen ausgelegt sind, ist die Anzahl der Hersteller und Anbieter sehr hoch. Die Angebote unterscheiden sich bezüglich Art, Leistung und Umfang nicht unerheblich.

Grundsätzlich gibt es zwei Arten von ERP Lösungen:

Einerseits gibt es die **branchenübergreifenden ERP-Lösungen**, die die meisten Funktionsbereiche und Geschäftsprozesse innerhalb eines Unternehmens abdecken und von daher weit weniger komplex und aufwendig sind als branchenindividuelle Systeme. Eingesetzt werden sie vor allem bei KMU.

Andererseits gibt es die **branchenspezifischen ERP-Softwarelösungen**. Diese Spezialisten kommen zum Beispiel verstärkt beim Handel, bei den Finanzinstituten und Versicherungen und ganz besonders in der Industrie zum Einsatz. Hier sprechen wir von hochkomplexen Lösungen, die in der Lage sind, Prozesse sehr exakt abzubilden und auch eine ganze Kette (z. B. die Lieferkette vom Lieferanten bis zum Kunden) wiederzugeben. Bei den branchenspezifischen ERP-Lösungen ist zudem ein Trend zu unternehmensübergreifenden ERP-Lösungen festzustellen. Sie werden nicht mehr, wie in der Vergangenheit oft üblich, nur unternehmensintern eingesetzt, sondern vernetzen sich über die ganze Lieferkette hinweg. Das heisst, es werden nicht mehr nur die direkten Zulieferer, sondern auch deren Lieferanten in das System eingebunden. So lassen sich Engpässe oder veränderte Produktions- und Lieferbedingungen schon lange im Voraus erkennen und bei der Planung berücksichtigen.

6.6 ERP-Hersteller

Der ERP-Markt für Grossunternehmen und Konzerne wird ganz klar von nur wenigen Unternehmen dominiert:

- SAP
- Oracle
- Microsoft

Der Markt für mittlere Unternehmen hingegen verteilt sich auf eine weitaus grössere Anzahl an Anbietern. Dazu gehören unter anderem:

- Abacus
- Sage
- Proffix
- IFS

6.7 Kosten eines ERP-Systems

Wenn ein Unternehmen mit dem Gedanken spielt, ein ERP-System einzuführen, geht es in der Regel nicht lange, bis das Traktandum «Kosten» zur Diskussion gestellt wird. Um die Kosten für eine ERP-Einführung abschätzen zu können, muss man verschiedene Faktoren berücksichtigen. Dies wären zum Beispiel:

- Grösse des Unternehmens
- Anforderungen an die ERP-Software
- Betriebsmodell

Neben den externen Kosten für ein ERP-System sollte man aber unbedingt auch die entsprechenden Kosten für

- Beratungsleistungen
- Schulungen
- Wartungen des ERP-Systems

berücksichtigen. Diese können nicht unerheblich sein. Um ein finanzielles Übernehmen im Hinblick auf die ERP-Systemauswahl und -einführung zu vermeiden, sollten sowohl die kurz- als auch die langfristigen Kosten eingeplant werden. Leider gibt es keine allgemeingültigen Richtwerte für die Kosten. Erfahrungswerte geben zumindest grobe Schätzungen an, mit denen man rechnen muss. Bei grossen Unternehmen können die Kosten eines solchen ERP-Projekts über eine halbe Million Franken betragen (bei Grosskonzernen kann der Betrag auch schnell mal achtstellig sein), mittlere Unternehmen sollten ein Budget von mehreren Hunderttausend Franken einplanen, während kleine Unternehmen Kosten im fünfstelligen Bereich erwartet.

Die Alternative zu solch hohen Investitionskosten gerade für kleine und mittlere Unternehmen stellen die Software-Mietmodelle (Software-as-a-Service, SAAS) dar, die häufig von Cloud-basierten ERP-Software-Anbietern angeboten werden. Bei dieser Form der Bereitstellung entfallen die hohen Einführungskosten und der monatliche Betrag für kleinere Unternehmen/Nutzer ist relativ überschaubar. Gleichzeitig kann das Unternehmen mit der Software wachsen und das Wachstum in der Software skalieren.

6.8 Aufbau des ERP-Systems / ERD

Moderne ERP-Systeme sind alle nach dem gleichen Konzept aufgebaut. Mithilfe relationaler Verbindungen werden Tabellen, Daten und Informationen miteinander verbunden. Die Visualisierung des aus der Verknüpfung

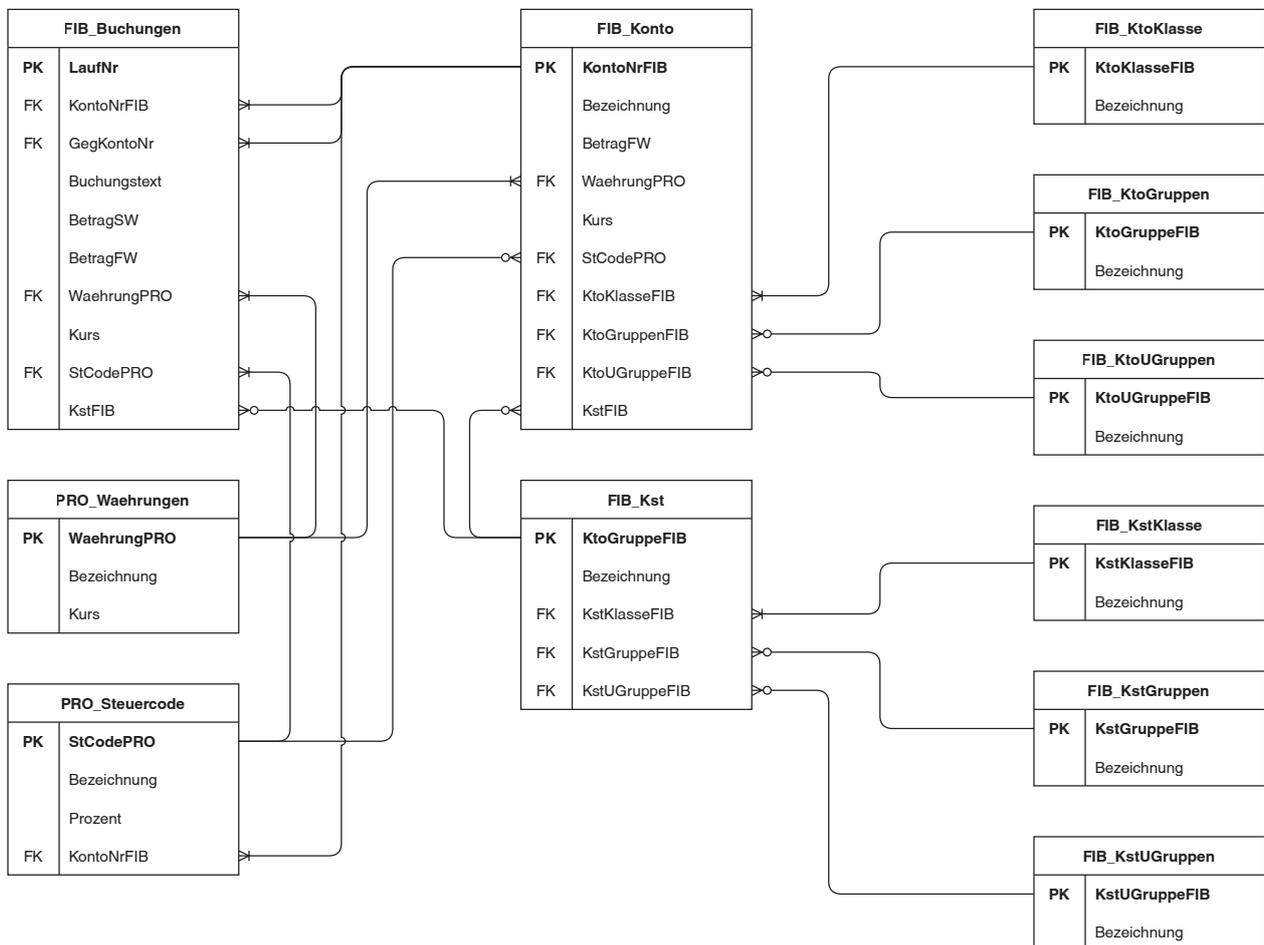


Abbildung 8: ERD-Diagramm (Proffix Px5 Finanzbuchhaltung, 2022)

entstandenen Netzwerkes von Tabellen und Daten wird **ERD-Diagramm** (Entity-Relationship-Diagramm) genannt und bildet sämtliche im ERP-System gespeicherten Informationen ab.

Dabei kann ein ERP-System rasch viele Tabellen umfassen, wie Abbildung 8 zeigt. Dem optimalen Aufbau der Datenbank kommt eine sehr hohe Bedeutung zu, indem mit effizienten Strukturen redundante Daten vermieden und eine hohe Zugriffsperformance auf die Daten erzielt werden kann.

Verarbeitungsstruktur in ERP-Systemen

Die Anzeige und die Verarbeitung von Informationen in ERP-Systemen erfolgen auf verschiedenen Ebenen. Für den versierten Benutzer und konzeptionell arbeitende Experten ist es wichtig, die Funktionen und Unterschiede zwischen diesen Ebenen zu kennen, da dies bei Problemsuche und Weiterentwicklung der Software entscheidend ist. Mitarbeitende benötigen heute kein vertieftes Verständnis der Prozesse mehr. Die hohe Prozessintegration führt dazu, dass sich Mitarbeitende nicht mehr mit Soll und Haben sowie Konten oder gar Mehrwertsteuer-Buchungen befassen müssen. Dies wird durch unterschiedlich eingerichtete Buchungsgruppen bewerkstelligt und im Hintergrund automatisch richtig gebucht (sofern die Stammdaten und Parameter richtig eingerichtet wurden).

Entscheidend für den versierten Nutzer und sein Verständnis ist es, Eingriffsebenen auf Softwareebene zu kennen und bei Anpassungen schlechte und gute Realisierungsansätze unterscheiden zu können.

Der grosse Teil der Literatur bezeichnet die Dateneingabe als User-Interface oder GUI. Diese werden in Formulare oder Journale geschrieben (1. und 2. Ebene der nachfolgenden Darstellung). Bei Buchungsprozessen werden diese Daten verändert (berechnet, kopiert, mutiert) und in neue Tabellen geschrieben. Im Falle von Lohnbuchungen sind diese Daten in den meisten Fällen nicht mehr mutier- oder veränderbar. Deshalb ist es wichtig, bereits vor den definitiven Buchungsprozessen prüfen zu können (auf Ebene 3), was am Ende gebucht wird, sowie gleichzeitig dafür zu sorgen, dass die eingegebenen und die gebuchten Daten wie Kosten, Termine, Material und Dimensionen (KST, KTR etc.) überprüf- und referenzierbar sind.

Jede Software funktioniert so, dass zwischen einzelnen Ebenen Informationen ausgetauscht werden. Normalerweise werden Informationen von oben nach unten übergeben, wobei Daten rekursiv, das heisst zwischen den einzelnen Ebenen, meistens nicht mehr verändert werden können. Alle ERP-Systeme haben gemeinsam, dass nach dem Buchen (3. auf 4. Ebene) die eingegebenen Informationen in der 3. Ebene (meistens) gelöscht werden und in der 4. Ebene unveränderbar (gebucht) vorliegen. Sollen Informationen in der 4. Ebene gelöscht werden, ist dies nicht oder nur mit grösseren Schwierigkeiten machbar, weil die Daten unter Umständen in mehrere Tabellen geschrieben wurden (z. B. von der Lohnbuchhaltung als Sammelbuchungen in die Finanzbuchhaltung und gleichzeitig in die Kostenrechnung).

Je nach System können über Storno-Funktionen Daten aus der Ebene 4 wieder auf die Ebene 3 zurückübertragen werden. Aufgrund von Sammelbuchungen wird jeweils eine Reihe von Datensätzen zurückgeschrieben resp. storniert. Dabei werden die veränderten Datensätze nicht gelöscht, sondern als gelöscht gekennzeichnet oder storniert (d. h. als Gegenbuchung mit umgekehrten Vorzeichen).

Es ist daher entscheidend, bereits vor dem Buchen auf der Ebene 3 sämtliche Informationen prüfen zu können, die danach auf die Ebene 4 gebucht werden. Aus diesem Grund gibt es sogenannte Testberichte oder Protokolle, um die Buchungen zu prüfen und beispielsweise FIBU-Konten abzustimmen.

Für die Entwicklung neuer Funktionen ist zu berücksichtigen, dass auf der richtigen Ebene in die Software eingegriffen wird und veränderte Buchungsprozesse reproduzierbar sind. Aus diesem Grund sind ERP-Hersteller vermehrt dazu übergegangen, Teile der Software vor Veränderungen zu schützen, damit Entwicklungsfehler keine substantziellen Auswirkungen haben. Dabei geht es vor allem um Buchungen ins FIBU-Hauptbuch sowie Materialbuchungen ins Lagerbuch.



Abbildung 9: Software-Verarbeitungsebenen (Zäch, 2021)

Aus diesem Grund sollte für Veränderungen an Buchungsprozessen immer Expertenwissen und -erfahrung konsultiert werden.

Beispiel Buchungsprozess von der Debitorenbuchhaltung in die FIBU

Die Abbildung 10 zeigt die Beziehungen zwischen den wichtigsten Tabellen und Verarbeitungen in der Debitorenbuchhaltung.

- In den Stammdaten werden allgemeine Informationen gespeichert (z. B. Adressinformationen, Währung, Geschäftsbuchungsgruppen, Debitorenbuchungsgruppen, Mehrwertsteuerbuchungsgruppen, Dimensionen wie Preisgruppen/Kostenstellen und -träger etc.).
- Die Buchungsgruppen wiederum greifen auf die Buchungsmatrizen zu, welche wiederum auf die FIBU-Konten referenzieren.
- In der Tabelle **Verkaufsauftragkopf** werden die Informationen des Kunden für den aktuellen Auftrag übertragen (meist als Kopie der Stammdaten).
- In der Tabelle **Verkaufsauftragzeilen** werden dem Kunden die **Buchungsgruppen des Artikels** zugeordnet (meist als Kopie der Stammdaten).
- Die Funktion Buchen (Lieferung und/oder Fakturierung) führt die **Verbuchung** durch und überträgt **den Auftrag** in die **gebuchten Lieferungen und die gebuchten Rechnungen**. Dabei werden die FIBU-Buchungen (z. B. Lagerabgänge, Mehrwertsteuerbeiträge und Rabatte) berechnet und verbucht. Eine Lieferung mit gebuchter Rechnung kann bei integriertem System nicht mehr rückgängig gemacht, sondern muss über Gutschrift ausgeglichen werden.
- Die meisten **Berichte** in der Debitorenbuchhaltung basieren auf den **gebuchten Lieferungen und Rechnungen**.

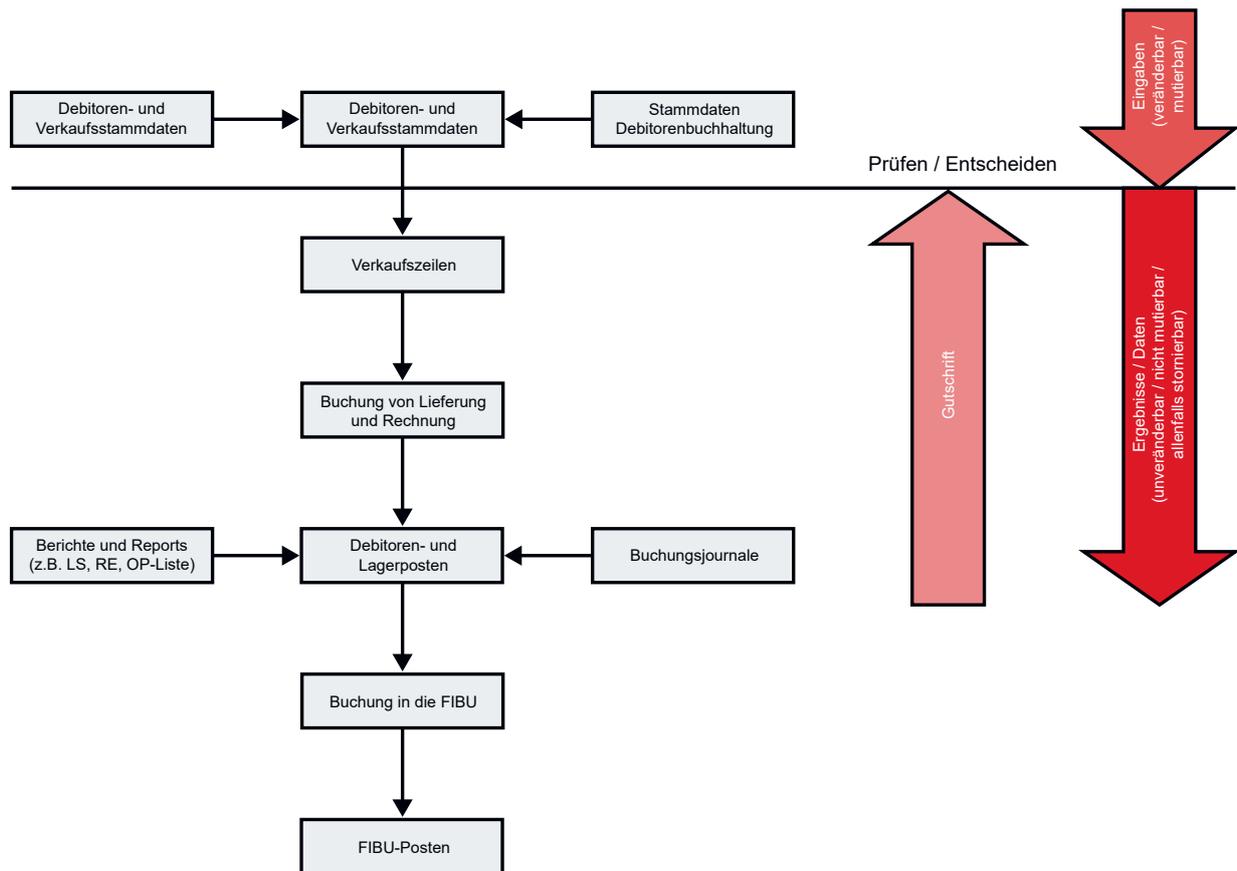


Abbildung 10: Aufbau Debitorenbuchungen (Zäch, 2021)

- Jede Lieferung und Rechnung wird im FIBU-Journal und weiteren betroffenen Nebenbüchern (Lager, MWST, Debitoren) dokumentiert.
- Je nach ERP-System kann definiert werden, ob die Daten aus dem Verkaufsmodul mit dem Verbuchen der **Debitoren- und Lagerposten** automatisch als **FIBU-Posten** in die FIBU gebucht werden. In diesem Fall wird von einem integrierten System gesprochen.

6.9 Grenzen und Perspektiven moderner ERP-Systeme

Die Grenzen heutiger ERP-Systeme liegen darin, dass sie den Blick vorwiegend nach innen, das heisst in das Unternehmen richten und damit symbolisch einen unternehmensinternen Kosmos abbilden. Dieser Blick zeigt gleichzeitig eine Momentaufnahme des Unternehmens, deren Nachteil darin liegt, dass retrospektiv bestimmte Situationen nicht reproduziert werden können (z. B. Bedarfsberechnungen logistischer Mengen, Mengenunterschreitungen, Kapazitätsbedarf an einem Stichtag oder Zustände vor dem Buchungsprozess). Der Blick nach innen und die Momentaufnahme ist in der globalisierten Welt nicht mehr zeitgemäss.

Der Blick nach aussen wird im besten Fall unidirektional über Schnittstellen zu geschäftsrelevanten Prozessen wie Zahlungsverkehr nach ISO 20022 (QR-Rechnung, SEPA), EDI/ELM/Zollabwicklung sichergestellt. Da die Anforderungen der Unternehmen im Blick nach aussen sehr unterschiedlich sind, kommt der Entwicklung von Schnittstellen eine grosse Bedeutung zu. Die Softwarehersteller bemühen sich daher zunehmend, über Datenbankschnittstellen im ERP-System den Blick nach aussen ebenfalls abzubilden und Schnittstellen reproduzierbar und transparent zu gestalten.

Im Zentrum der Schnittstellen stehen dabei mehrere Perspektiven:

1. Offene Datenbankschnittstellen und intelligente Informationszuordnung durch die Softwarehersteller. Vorbereitete Anbindung von Konkurrenz- oder Drittsystemen
2. Einbindung externer Informationen im ERP-System (z. B. demografische, gesellschaftliche oder konjunkturelle Daten)
3. Schaffung konfigurier- und personalisierbarer Visualisierungen in Dashboards/Cockpits, um den Mitarbeitenden nach Aufgabengebiet und Funktion eine personalisierte Übersicht über Aufgabenschwergewichte und Prioritäten zu geben
4. Bidirektionale Automatisierung des Datenaustauschs mit Geschäftspartnern zur Effizienzsteigerung sowie der Vermeidung redundanter Daten und Verbesserung sowie laufender Kontrolle der Datenqualität
5. Export von Informationen für interne und externe Zwecke (z. B. interne Auswertungen oder Meldungen an Partner wie Behörden, Versicherungen Verbände)

Die Weiterentwicklung der ERP-Systeme ist nie abgeschlossen, weil sich die unternehmerischen Bedürfnisse ebenfalls laufend weiterentwickeln. Gleichzeitig mit dem Blick nach aussen, der Globalisierung und zunehmender Unsicherheiten in der VUCA-Welt nimmt das Bedürfnis zu, Informationen zu vergleichen, Trends zu erkennen und im besten Fall durch aussagekräftige Vorhersagen (Predictions) zeitoptimal die unternehmerisch richtigen Massnahmen einleiten zu können.

Während die Verwaltung von Daten und Informationen durch die Normierung von Schnittstellen und Tools auch für den Anwender vereinfacht wird, werden die Datenstrukturen der Organisationen komplexer, während die zu verwaltenden Datenmengen exponentiell wachsen. Die Überwachung sowie die Verwaltung von Daten sowie deren Qualität erhalten in Zukunft eine höhere Priorität und stellen Organisationen vor neue Herausforderungen.

In diesen Organisationen werden sich die Aufgaben des Controllers durch die gesamtheitliche Auswertung und Analyse von Daten, die Weiterentwicklung durch neue Controllingansätze und -perspektiven substanziell verändern. Der Controller wird zum Datenanalysten operativer und strategischer Informationen und zum Steuermann der operativen und strategischen Unternehmensentwicklung.

7 Business Intelligence (BI)

7.1 Definition von Business Intelligence

Business Intelligence ist ein technologiegetriebener Prozess zur Analyse von Daten und zur Präsentation verwertbarer Informationen, der Führungskräften, Managern und anderen Endanwendern hilft, fundierte Geschäftsentscheidungen zu treffen.

Business Intelligence als Prozess umfasst sämtliche Formen der systematischen Sammlung, Auswertung, Analyse und Darstellung von Informationen. Dabei steht immer die Frage im Vordergrund: Wie können Unternehmen grosse Datenmengen verwalten und für eine bessere Entscheidungsfindung nutzen?

Datenmengen sind in diesem Zusammenhang im engeren Sinne immer als Informationen zu verstehen, die für eine Entscheidungsfindung herangezogen werden.

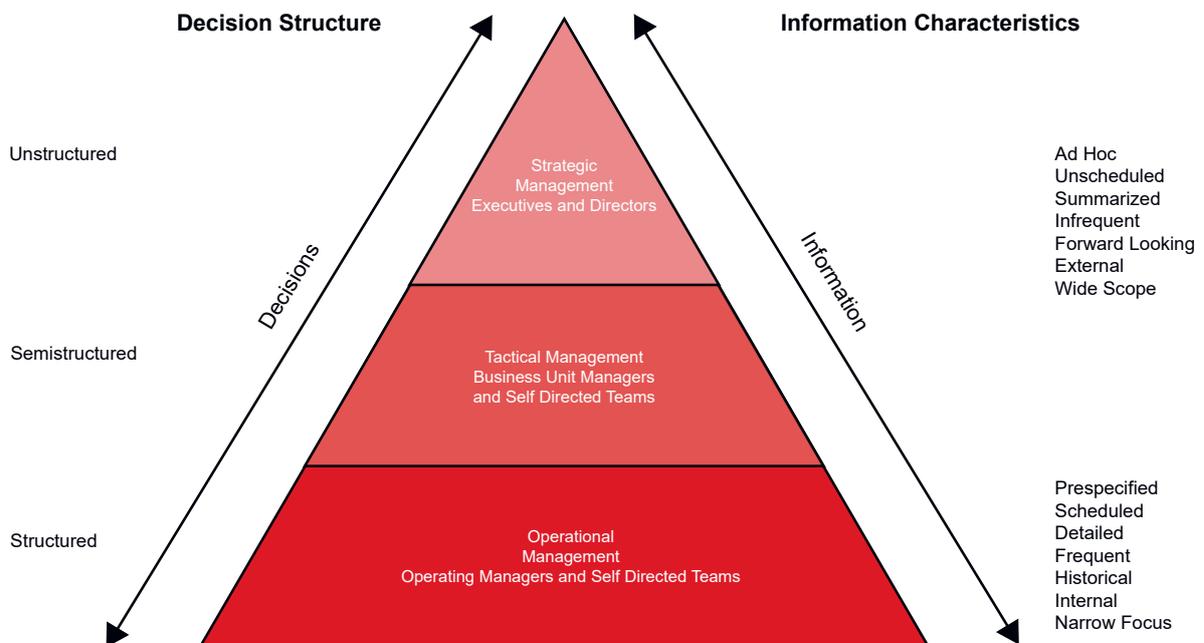


Abbildung 11: Entscheidungsebene und Information (O'Brien, 2010)

Betrachtet man die obige Abbildung, ist ein BI vor allem für das strategische Management von Bedeutung. Dies ist im Kontext der Informationsverarbeitung und der daraus abgeleiteten Entscheidung zu sehen. Verdeutlichen lässt sich dies mit folgendem **Beispiel**: Das operative Management bekommt die sehr strukturierte Information, dass ein Kunde mit seiner Zahlung im Verzug ist. Die daraus abzuleitende Entscheidung ist relativ einfach: Der Kunde sollte eine Zahlungserinnerung erhalten. Für eine solche Entscheidung ist kein BI notwendig, da die Information dafür sehr strukturiert vorhanden ist.

Anders sieht es mit strategischen Entscheiden aus, hier sind die Informationen meistens sehr unstrukturiert, und eine Entscheidung daraus abzuleiten, ist um ein Vielfaches komplexer. Hierzu ein weiteres Beispiel:

Unternehmen versinken heute in Daten. Sie haben Daten aus Kundenprofilen, ERPs, Loyalty-Programmen, Marktanalysen usw. Es werden mehr Daten gespeichert, als der Mensch in der Lage ist, zu analysieren. Zusätzlich haben wir den Trend, dass die Speicherkapazitäten weiterhin steigen und die Kosten für die

Speicherkapazität weiter sinken. Was bedeutet diese Entwicklung für die BI-Branche? Fazit: Sobald Informationen unstrukturierter werden, ist eine Entscheidungsfindung deutlich komplexer.

Wenden wir uns noch kurz der historischen Entwicklung zu. Der Begriff «BI» wird erstmals im 1958 veröffentlichten Beitrag «A Business Intelligence System» von Hans Peter Luhn verwendet – etabliert und geprägt hat ihn dann in den frühen 1990er-Jahren die Gartner Group.

Business Intelligence entwickelte sich zunächst relativ unabhängig von Data Warehousing, Enterprise Content Management und Knowledge Management. Business-Intelligence-Systeme werden vor allem im deutschsprachigen Raum als analytische Informationssysteme verstanden. Inzwischen haben Data Warehousing und Business-Intelligence-Systeme ein gewaltiges Wachstum, eine zunehmende Bedeutung für das Informationsmanagement sowie einige Paradigmenwechsel und Erweiterungen erfahren.

7.2 Warum ist Business Intelligence wichtig?

Ursprünglich wurden BI-Tools hauptsächlich von Datenanalysten und anderen IT-Experten verwendet, die Analysen durchführten und Berichte mit Abfrageergebnissen für Geschäftsanwender erstellten. Klassische Einsatzgebiete sind vor allem die Unternehmensbereiche Finanzen und Controlling. In zunehmendem Masse nutzen jedoch Führungskräfte und Mitarbeiter BI-Plattformen selbst, was zum Teil auf die Entwicklung von Selfservice-BI, Datenerfassungs-Tools und -Dashboards zurückzuführen ist.

Um den immer grösser werdenden Datenmengen Herr zu werden, haben sich in den vergangenen Jahren IT-gestützte Business-Intelligence-Systeme in den Unternehmen durchgesetzt.

Gerade im Zusammenhang mit Big Data wird die analytische Komponente immer wichtiger (tableau.com, 2021). Dazu kommt der Paradigmenwechsel – vom automatischen System zur Verbreitung von Informationen an die verschiedenen Abteilungen von Industrie-, Wissenschafts- und Regierungsorganisationen hin zum Business Performance Measurement.

7.3 Warum Business Intelligence nutzen?

Zu den potenziellen Vorteilen von Business-Intelligence-Tools gehören die Beschleunigung und Verbesserung von Entscheidungsprozessen, die Optimierung interner Geschäftsprozesse, die Steigerung der betrieblichen

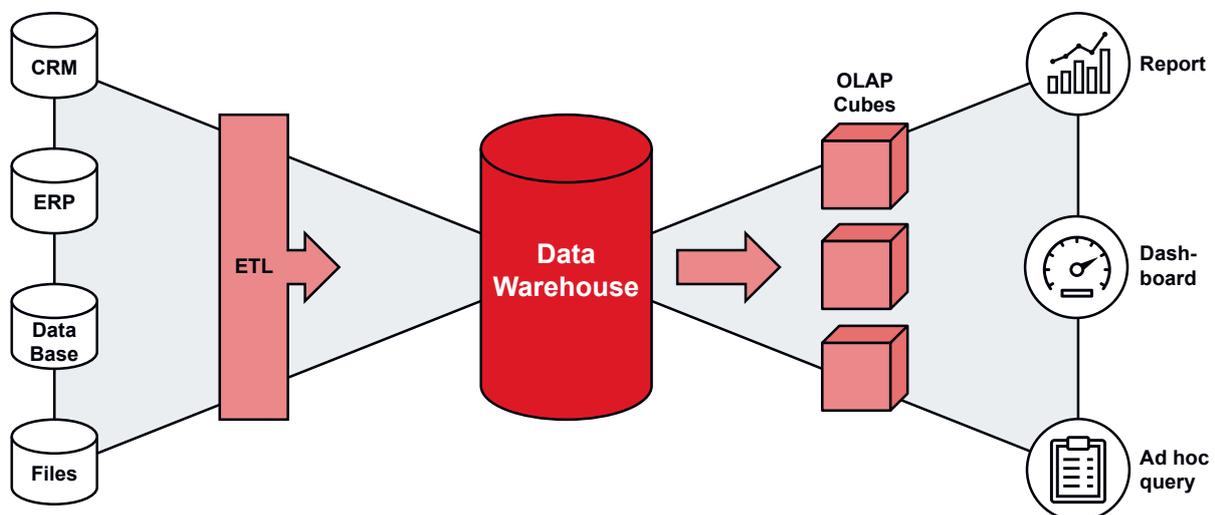


Abbildung 12: BI back/Frontend

Effizienz, die Steigerung neuer Umsätze und der Gewinn von Wettbewerbsvorteilen gegenüber Mitbewerbern.

Wenn wir Abbildung 12 einer BI-Architektur betrachten, sehen wir, dass das BI auf multiple Quellen zugreift (von links her betrachtet), diese in einem Data Warehouse zusammenführt und dann in verschiedenen, thematisch abgegrenzten OLAP-Cubes aufbereitet und die Daten anschliessend via verschiedenen Frontends ausgibt. Der grosse Vorteil gegenüber einem reinem ERP ist hierbei, dass man nicht nur auf statische Daten zugreifen kann. Es gibt uns die Möglichkeit, innovative Wege zu gehen und auch Daten von Loyalty-Programmen, Social Networks usw. zu sammeln, diese strukturiert aufzubereiten, damit durch Mustererkennung vorausschauende Entscheidungen zu treffen und nicht permanent reaktiv zu entscheiden. Die Unternehmung gelangt damit in die Situation zu wissen, was der Kunde in Zukunft haben möchte.

7.4 Business-Intelligence-Software

Business-Intelligence-Lösungen verfügen über unzählige Funktionen, aber im Allgemeinen kann man damit alle oder einen Teil der folgenden Aufgaben durchführen:

- Vergleich, Abgleich verschiedener Ergebnisse über A/B- oder multivariate Tests
- Prognostizieren von Ergebnissen durch prädiktive Modellierung
- Einblicke in Muster oder Beziehungen
- Präsentieren von Ursache-Wirkungs-Szenarien

Legen Sie beim Vergleich von Business-Intelligence-Lösungen Wert darauf, dass sie alle oder zumindest einige der folgenden Funktionen enthalten sollten:

7.5 Online Analytical Processing (OLAP) / Business Analytics

OLAP umfasst die gesamte Bandbreite der Datenaufbereitung, -verarbeitung und -bereitstellung für das Reporting. Den Kern der BI-Lösungen bilden OLAP-Tools, mit denen Sie multidimensionale Daten in Echtzeit oder gleichzeitig mit anderen Analyseverfahren analysieren können. Generell gilt: Je mehr Möglichkeiten Sie haben, Daten aus einem Dataset (dem sogenannten OLAP-Cube) zu isolieren und anzuzeigen, desto flexibler können Sie Daten interpretieren.

Business Analytics ist ein Prozess der Datenveredelung und Datenauswertung zur strategischen Unternehmenssteuerung. Es ist ein digitaler Prozess, in dem nicht nur Massendaten gesammelt, sondern auch durch statistische und maschinelle Methoden angereichert, aufbereitet und ausgewertet werden.

Business Analytics sammelt und analysiert Daten wie Business Intelligence, erkennt Datenkorrelationen und Muster bei zunehmenden Datenmengen (Big Data) mithilfe von Algorithmen und führt vor allem prädiktive Analysen durch, um die Wahrscheinlichkeit künftiger Ergebnisse zu ermitteln. Wie bei Business Intelligence werden diese Berichte in benutzerdefinierten Dashboards visualisiert.

7.6 Reporting und Abfragen

Reporting und Abfrage ist ein weiterer Schwerpunkt von BI-Software, die komplexe Daten generiert und sortiert. Inzwischen verfügen moderne BI-Lösungen über intuitive Abfragen, die es Anwendern ermöglichen, komplexe Daten auch ohne SQL-Schreib- oder Programmierkenntnisse abzubilden. OLAP bereitet die Datensätze automatisch für jede neue Abfrage vor. Dies ermöglicht es Unternehmen, BI-Tools unternehmensweit und über den traditionellen Bereich der IT-Abteilung hinaus einzusetzen.

7.7 Digitale oder visuelle Dashboards

Die Datenvisualisierung ist die Endphase des Berichtswesens und tatsächlich eines der identifizierbarsten Elemente von BI-Lösungen. Führungskräfte und nichttechnische Benutzer betrachten BI oft ausschliesslich aus diesem Blickwinkel. Erweiterte Visualisierungen umfassen raumbezogene Berichte, Wärmekarten etc.

7.8 Integration

BI-Lösungen basieren auf verschiedenen Datenquellen, um aussagekräftige Erkenntnisse zu gewinnen, weshalb die Integration von entscheidender Bedeutung ist. Tatsächlich bündeln viele Anbieter BI-Tools mit anderen Geschäftsanwendungen wie CRM, Salesforce-Automatisierung und Warenkorb-Software.

8 Big Data

Wenn wir von Big Data sprechen, meinen wir Datenmengen, die so gross sind, dass man sie mit normaler Software und Hardware nicht mehr verarbeiten kann. Wie entstehen nun Datenmengen, die so gross sind? Die heutige Zeit zeichnet sich durch eine wortwörtliche Datensammelwut aus. Heutzutage wird alles, was wir machen und tun, irgendwo gespeichert. Sie kaufen beispielsweise online ein oder suchen etwas im Internet, Sie geben eine Route in Ihr Navigationssystem ein oder machen eine Finanztransaktion, Sie telefonieren oder gehen ins Fitnessstudio – egal was, alles wird gespeichert. Dies führt zu einem gewaltigen Datenvolumen, das auch genutzt werden will. Man schätzt, dass dieselbe Datenmenge, die die Menschheit seit ihrer Entstehung bis zum Jahr 2002 gesammelt hat, im Jahr 2014 in zehn Minuten entstanden ist und sich nun alle zwei Jahre verdoppelt.

Mit Big Data ist aber nicht nur die Datenmenge an sich gemeint, sondern vor allem auch deren Analyse, um Muster daraus zu erkennen oder abzuleiten. Prominentes Beispiel hierfür, was man heute beinahe täglich erlebt, ist das Onlineshopping. Sie entscheiden sich, den Artikel X zu erwerben. Kaum haben Sie diesen zu Ihrem Warenkorb hinzugefügt, sehen Sie sofort die Übersicht, die Ihnen zeigt, was andere Kunden, die denselben Artikel erworben haben, sonst noch gekauft haben (und Ihnen daher vielleicht auch gefallen könnte?). Solche Informationen werden in Echtzeit berechnet und geben sogar die Höhe der Wahrscheinlichkeit an, dass Ihnen dieser Artikel auch gefallen könnte.

8.1 Wo wird Big Data genutzt?

- **Medizinische Forschung:** Durch die Auswertung von Massendaten können Mediziner die besten Therapielösungen und -pläne für ihre Patienten finden.
- **Industrie:** Dank der Nutzung eigener Maschinendaten können Unternehmen die Effizienz ihrer Produktion steigern und so nachhaltiger arbeiten.
- **Wirtschaft:** Big Data ermöglicht es Unternehmen, ihre Kunden besser kennenzulernen und Angebote besser auf diese abzustimmen.
- **Energie:** Um den Energieverbrauch auf individuelle Bedürfnisse zuzuschneiden, muss man diese Bedürfnisse kennen. Erhobene Verbrauchsdaten sorgen langfristig für eine nachhaltige Energieversorgung.
- **Marketing:** Big Data wird im Marketingbereich oft zu Targeting-Zwecken genutzt. Ziel ist meist die Verbesserung der Customer Relationship und eine Steigerung von Conversions bei verschiedenen Marketingmassnahmen.
- **Verbrechensbekämpfung:** Auch Regierung und Staatsschutz greifen auf Big Data zurück, z. B. im Rahmen der Terrorismusbekämpfung.
- **Politik:** Big Data wird auch im Wahlkampf eingesetzt, um potenzielle Neuwähler zu gewinnen, Ergebnisse vorherzusagen und allenfalls zu beeinflussen.

8.2 Big Data auswerten

Hier können wir den Bogen zum vorhergehenden Kapitel schlagen. Business Intelligence ist der Prozess, mit dem sich Big Data aufnehmen, analysieren und anwenden lässt. Die oben beschriebenen Vorteile sind durch die systematische Auswertung von Big Data in BI-Lösungen entstanden und bilden damit ein wichtiges Instrument im Kampf um Marktanteile. Durch die Darstellung und Vorhersage von Chancen und Herausforderungen können Organisationen mit Business Intelligence ihre Big Data optimal für ihren Erfolg nutzen.

8.3 Big Data nutzen

Wie geht man als Unternehmen vor, wenn man Big Data nutzen möchte? Oberstes Gebot ist hier sicher eine gute Planung. Aber welche Elemente muss man in der Planung von Big Data berücksichtigen? Experten sprechen hier häufig von den **5 V**:

- **Volume**: Planen Sie das Handling der vorgesehenen Datenmenge. Wie möchten Sie diese Daten speichern und – ganz wichtig – wo?
- **Variety**: Identifizieren Sie die Standorte sämtlicher Datenquellen und evaluieren Sie die besten Tools, um die Daten zu schürfen resp. zu erfassen.
- **Velocity**: Besorgen Sie sich die richtige Technologie, um Ihre Daten auszuwerten und damit aussagekräftige Analysen erstellen zu können.
- **Veracity**: Stellen Sie unbedingt sicher, dass Ihre Daten sauber und korrekt sind.
- **Value**: Designen Sie eine Big-Data-Umgebung, mit der Sie aussagekräftige Analysen erstellen können und Wichtiges von Unwichtigem unterscheiden. Nicht alle gesammelten Daten sind wichtig für Ihr Unternehmen.

8.4 Wo speichert man grosse Datenmengen?

Grundsätzlich bieten sich hier zwei Möglichkeiten an:

Data Warehouse

In einem Data Warehouse lassen sich grosse Informationsmengen abspeichern und für eine weitere Auswertung bereitstellen. Das Data Warehouse führt verschiedene Daten in einheitlichen Formaten und Strukturen zusammen. Diese richten sich danach, was für eine Analyse durchgeführt werden soll. Das bedeutet, in einem Data Warehouse befinden sich ausschliesslich aggregierte Daten wie Kennzahlen oder Transaktionsdaten. Es lassen sich nur Daten abspeichern, die bereits verarbeitet wurden und einem vorab bestimmten Zweck dienen. Es ist schwer, die darin abgelegten Informationen zu verändern.

Data Lake

Dabei handelt es sich um ein zentrales Speicherrepository, das Big Data aus vielen Quellen in einem rohen, granularen Format enthält. Es kann strukturierte, semistrukturierte oder unstrukturierte Daten speichern. Das heisst, die Daten können in einem flexibleren Format zur späteren Nutzung aufbewahrt werden. Ein Data Lake verbindet Daten beim Speichern mit Identifiern und Metadaten-Tags für einen schnelleren Zugriff. Data Scientists können mit Data Lakes schneller und mit einer höheren Genauigkeit Daten abrufen, vorbereiten und analysieren. Analyse-Experten können aus diesem Datenpool nach Bedarf Daten für verschiedenen Use Cases wie Sentimentanalysen oder zur Betrugserkennung abrufen.

8.5 Big-Data-Softwarelösungen

Zu den bekanntesten Tools gehören sicher:

- Hadoop: Open-Source-Softwarelösung für Big Data (Apache Software Projects)
- MapReduce (Apache Software Projects)
- Spark: Open-Source-Softwarelösung für Big Data (Apache Software Projects)

8.6 Kritik an Big Data

Bei allen Vorteilen von Big Data werden aber auch kritische Fragen laut:

- Wem gehören diese Daten?
- Wollen wir, dass nur wenige Unternehmen diese Daten kontrollieren? Prominentes Beispiel ist hier Google. Google hat eine enorme Datenhoheit. Hier wird die Monopolstellung wie auch die Gefahr eines Datenmissbrauchs kritisiert.
- Wollen wir, dass menschliches Verhalten vorhergesagt werden kann?
- Was passiert, wenn diese Daten in die falschen Hände geraten?

Man kann Big Data in diesem Kontext als Fluch oder als Segen betrachten. Dies wird primär davon abhängig sein, wie diese Daten in Zukunft genutzt werden. Bei aller Kritik muss jedoch auch gesagt werden, dass Big Data sehr wohl sinnvoll sein kann. So wurden zum Beispiel in der Krebsforschung grosse Fortschritte erzielt, die ohne Big Data nicht möglich gewesen wären. Ebenso wird die Energieversorgung und das Verkehrswesen durch die Auswertung von Massendaten stetig optimiert und weiterentwickelt, was der gesamten Bevölkerung zugutekommt. Sicher ist auf jeden Fall nichtsdestotrotz, dass unsere Zukunft von grossen Datenmengen geprägt sein wird.

9 SQL

9.1 Einführung

Die Abkürzung SQL steht für Structured Query Language und ist eine Datenbanksprache zur Erstellung von Datenbankstrukturen in relationalen Datenbanken sowie zum Bearbeiten und Abfragen der darauf basierenden Datenbestände. SQL ist in Englisch, resp. stark an die englische Sprache angelehnt.

Mit SQL können Abfragen erstellt werden, die Daten in relationalen Datenbanken abrufen und dem Benutzer mittels einer Anwendersoftware resp. einem Frontend bereitstellen.

SQL ist heute der gängige Standard, um Abfragen auf relationalen Datenbanken auszuführen. Neben SQL gibt es noch die Datenbanksprache MDX, die in multidimensionalen Datenbanken verwendet wird.

9.2 Befehl SELECT

Der SQL-Befehl SELECT wird in der Regel als Basis für SQL-Abfragen verwendet, die auf einer Datenbank ausgeführt werden sollen. Mithilfe dieses Befehls ist es möglich, Daten aus einer oder mehreren Tabellen über eine JOIN (Verbindung) abzufragen.

SQL SELECT Beispiel

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Für das SQL SELECT-Beispiel wird die Tabelle Kundenstammdaten verwendet. Möchte man von dieser Tabelle nur den Namen der Kunden sowie ihren Wohnort ausgeben, ist folgender SQL-Befehl einzugeben:

```
SELECT NAME, ORT FROM KUNDENSTAMMDATEN
```

Die Ergebnismenge der SQL Select-Abfrage würde folgendermassen aussehen:

Vorname	Ort
Monika	Niederglatt
Thomas	Zürich
Felix	Niederglatt
Regula	Zürich
Reto	Niederglatt
Stefano	Flaach
Barbara	Diesldorf

9.3 Befehl DISTINCT

Der SQL DISTINCT-Befehl wird in einer SQL Select-Abfrage direkt hinter dem Select platziert. Mithilfe des DISTINCT-Befehls wird vermieden, dass man im Abfrageergebnis doppelte Werte hat.

SQL DISTINCT Beispiel

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Mit der SQL-Abfrage Select Distinct können wir prüfen, ob bestimmte Spalten redundanzfrei sind – und falls nicht, diese ohne Redundanzen erstellen. In unserem Beispiel Kundenstammdaten ist die KD_ID ein Feld, das ohne Redundanzen sein sollte, da die KD ID pro Kunde einmalig sein soll (eine ID ist in der Regel immer ein Primärschlüssel und darf keine doppelten Werte haben).

Der Befehl, um obige Kundenstammdatenliste redundanzfrei zu erstellen, würde wie folgt aussehen: SELECT DISTINCT KD_ID FROM Kundenstammdaten

Die Ergebnismenge der SQL Select-Abfrage würde folgendermassen aussehen:

KD_ID
4721
5397
6423
8174
9371
7348

Natürlich wäre das Problem der doppelten Kunden-ID noch nicht gelöst. Man müsste auch einerseits eine Stammdatenbereinigung vornehmen und andererseits die Kundenstammdaten so aufbauen, dass gar keine doppelten Kunden-IDs möglich sind.

9.4 Befehl WHERE

Mit dem SQL WHERE-Befehl kann man Datensätze, eingeschränkt nach bestimmten Kriterien, auswählen. Der SQL WHERE-Befehl funktioniert eigentlich wie ein Filter, mit dem man Datensätze einschränken kann. Man kann den WHERE-Befehl mit folgenden Parametern spezifizieren:

- Gleich (=) oder ungleich (<>)
- Grösser als (>) oder kleiner als (<)
- Grösser oder gleich (>=), kleiner oder gleich (<=)

SQL WHERE Beispiel

Bezugnehmend auf unser Kundenstammdaten-Beispiel:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Wir möchten nur Datensätze (alle Spaltenüberschriften) haben, bei denen die PLZ = 8172 ist. Hierzu verwenden wir folgenden SQL-Befehl:

```
SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE PLZ = 8172
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.5 SQL-Operatoren AND / OR

Die SQL AND & OR-Operatoren werden in der Regel in SQL-Where-Bedingungen eingebaut, um bestimmte Abfrageergebnisse zu erzielen. Sie sind im weitesten Sinne mit Excel-Wenns-Funktionen zu vergleichen. Die AND & OR-Befehle basieren auf der booleschen Algebra. Mithilfe des AND-Operators werden zwei oder mehrere SQL-Bedingungen zusammengefasst. Der OR-Operator sorgt für eine Unterscheidung zwischen zwei oder mehreren SQL-Bedingungen. Ein OR-Operator wiegt immer schwerer als ein AND-Operator, das heisst, sollte ein Konflikt zwischen einem AND-Operator und einem OR-Operator entstehen, ist der OR-Operator immer mächtiger und wird den AND-Operator übersteuern.

SQL WHERE AND Beispiel

Wir starten wieder mit unserem Kundenstammdaten-Beispiel.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Wir möchten nur Datensätze (alle Spaltenüberschriften), bei denen die PLZ = 8172 ist und die Kunden männlich (Geschlecht = m). Hierzu verwenden wir folgenden SQL-Befehl:

SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE PLZ = 8172 AND Geschlecht = "m"

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

SQL WHERE OR Beispiel

Wir starten wieder mit unserem Kundenstammdaten-Beispiel.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Wir möchten nur Datensätze (alle Spaltenüberschriften), bei denen die PLZ = 8172 oder = 8000 ist. Hierzu verwenden wir folgenden SQL-Befehl:

SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE PLZ = 8000 OR PLZ = 8172

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.6 SQL IN Operator

Mit dem SQL IN Operator kann man in SQL-Where-Bedingungen mehrere Abfrageergebnisse einer SQL-Anweisung zusammenfassen. Somit kann man mehrere OR-Operatoren ersetzen und damit Bedingungsstrukturen deutlich vereinfachen.

SQL IN Operator Beispiel

Wir knüpfen hier an unserem Beispiel der Kundenstammdaten (WHERE-Bedingung) an und zeigen auf, wie man diese Problematik mit dem IN-Operator lösen kann.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Wir möchten nur Datensätze (alle Spaltenüberschriften), bei denen die PLZ = 8172 oder = 8000 ist. Hierzu verwenden wir folgenden SQL-Befehl:

```
SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE PLZ IN (8000, 8172)
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.7 SQL-Operator BETWEEN

Wenn Sie mit einer SQL-Abfrage einen Datenrange abbilden möchten, bietet sich der BETWEEN-Operator an. Mit diesem können Sie auf einfache Weise einen Datenrange ausgeben.

SQL-Operator BETWEEN Beispiel

Wir knüpfen hier wieder an unserem Beispiel der Kundenstammdaten an. Um über eine Datumsspalte zu verfügen, haben wir zusätzlich das Feld Kunde seit eingefügt.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Wir möchten nur Datensätze (alle Spaltenüberschriften), bei denen die Kunden im Range seit 2011 bis 2015 registriert wurden. Hierzu verwenden wir folgenden SQL-Befehl:

```
SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE Kunde_seit BETWEEN 2011 AND 2015
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15

9.8 Befehl SQL LIKE

Bis hierhin haben wir nur Befehle und Operatoren besprochen, die auf einem festen Wert beruhen resp. die ein Kriterium fest vorgeben (z. B. Geschlecht, PLZ, Kunde seit, etc.). Was machen wir aber, wenn wir Daten nach einem Fragment selektieren möchten, beispielsweise alle Kunden, bei denen der Vorname mit R beginnt? Hier können wir den SQL LIKE-Befehl verwenden.

Mit dem LIKE-Befehl können wir Daten nach folgenden Mustern durchsuchen (hier am Beispiel Namen aufgezeigt):

- "B_L": Alle Namen, die mit einem B beginnen und mit einem L enden.
- "BE*": Alle Namen, die mit BE beginnen (es gehen auch mehr als zwei Zeichen)
- "*LE": Alle Namen, die mit LE enden (es gehen auch mehr als zwei Zeichen)
- "**SP*": Alle Namen, die an irgendeiner Stelle die Zeichen SP enthalten (es gehen auch mehr als zwei Zeichen).

SQL LIKE Befehl Beispiel

Wir knüpfen hier wieder an unserem Beispiel der Kundenstammdaten an. Wir möchten aus unseren Kundenstammdaten alle Kunden, deren Name auf CH endet.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Um das zu erreichen, verwenden wir folgenden SQL-Befehl:

```
SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten WHERE Name LIKE "*CH"
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

9.9 SQL ORDER BY Operator

Bis hierhin haben wir uns vor allem auf Einschränkungen (ähnlich EXCEL-Wenns-Funktionen) konzentriert. Selbstverständlich kann man mit SQL aber auch Sortierungen vornehmen. Ähnlich wie in Excel hat man dabei auch die Möglichkeit, auf- bzw. absteigend zu sortieren. Der entsprechende SQL-Befehl hierzu ist:

- ASC: Aufsteigend sortieren
- DESC: Absteigend sortieren

SQL ORDER BY Beispiel

Wir möchten unsere Kundenstammdaten nach dem Feld KD ID absteigend sortieren:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Der entsprechende SQL-Befehl hierzu sieht folgendermassen aus:

```
SELECT Vorname, Name, Geschlecht, KD_ID, Strasse, Nr, PLZ, Ort, Kunde_seit, Transaktionen FROM Kundenstammdaten ORDER BY KD_ID DESC
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2

9.10 Befehl SQL GROUP BY

Möchten wir unsere Daten nach bestimmten Parametern gruppieren, können wir auf den SQL-Befehl GROUP BY zurückgreifen. Häufig wird dieser Befehl zusammen mit einer Aggregatfunktion ausgegeben. Aggregatfunktionen sind AVG, COUNT, MAX, MIN und SUM.

SQL GROUP BY Beispiel

Wir möchten von unseren Kundenstammdaten zum Beispiel wissen, wie viele Kunden wir seit 2014 neu haben.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Lorenz	Fischer	m	5401	Fischerweg	7	8000	Zürich	2014	14
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Sabine	Meier	w	6355	Meierstrasse	1	8812	Horgen	2015	4
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Der entsprechende SQL-Befehl hierzu sieht folgendermassen aus:

```
SELECT Kunde_seit, COUNT (Kunde_seit) as Anzahl_Kunden From Kundenstammdaten WHERE Kunde_seit >= 2014 GROUP BY Kunde_seit
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Kunde seit	Anzahl_Kunden
2014	2
2015	2
2016	1

9.11 Befehl SQL HAVING

Der SQL HAVING Befehl ist das SQL WHERE in einem SQL GROUP BY Befehl. Sie möchten beim vorhin besprochenen SQL GROUP BY Befehl noch eine zusätzliche Aggregation haben von einem weiteren Attribut? Genau das erreichen Sie mit SQL HAVING. Dieser Befehl kann die Ergebnismenge auf Basis der Aggregatfunktionen (AVG, COUNT, MAX, MIN, SUM) weiter einschränken und ausgeben.

SQL HAVING Beispiel

Sie möchten wieder auf Basis unserer Kundenstammdaten die Anzahl Kunden nach Kunde seit einschränken und wollen aber nur die Kunden selektiert haben, die seit Registrierung mindestens 15 Transaktionen getätigt haben.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Lorenz	Fischer	m	5401	Fischerweg	7	8000	Zürich	2014	14
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Sabine	Meier	w	6355	Meierstrasse	1	8812	Horgen	2015	4
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Der entsprechende SQL-Befehl hierzu sieht folgendermassen aus:

```
SELECT Kunde_seit, COUNT (Kunde_seit) as Anzahl_Kunden From Kundenstammdaten WHERE Kunde_seit >= 2014 GROUP BY Kunde_seit, Transaktionen HAVING Transaktionen >=15
```

Die Ergebnismenge der SQL-Abfrage sieht folgendermassen aus:

Kunde seit	Anzahl_Kunden
2014	1
2015	1

10 Robotic Process Automation (RPA)

10.1 Einleitung

Haben Sie bei Ihrem Arbeitstag wiederkehrende Tätigkeiten, die Sie als langweilig und gleichzeitig fordernd empfinden? Nicht selten erfordern monotone Arbeitsprozesse eine hohe Fokussierung, damit die Ergebnisse präzise genug sind. Der Mensch arbeitet ungern einen Arbeitsschritt zigmal durch und neigt daher dazu, seinen Fokus mit der Zeit zu verlieren.

Für Abstimmungen von Neben- und Hauptbüchern, Konzern- und Einzelabschlüssen und Geldflussrechnungen wird im Controlling viel manuelle Arbeitszeit investiert. Bei hohen Buchungsvolumen beschränkt sich daher oft die Abstimm- bzw. Kontrolltätigkeit auf Stichproben, da nicht jede Buchung oder jedes Konto abgeglichen werden kann.

Das Controlling wird sehr stark von den Zukunftstrends wie zum Beispiel Big Data gefordert. Neue Konzepte im Reporting sind sehr stark gefragt und daraus leiten sich viele Weiterentwicklungsprojekte ab. Gleichzeitig wird die Routinearbeit im unveränderten Ausmass erwartet. Hier braucht das Controlling eine Entlastung, um seine Ressourcen in die Weiterentwicklung investieren zu können.

Bei der Robotic Process Automation wird mithilfe von Software und teilweise auch Hardware die manuelle Tätigkeit automatisiert. Die Vorteile liegen auf der Hand: Das Controlling wird von Routinearbeit entlastet und der «Roboter» hat ein enormes Skalierungspotenzial. Im Gegensatz zum Menschen ist der Roboter in der Lage, einen Schritt x-fach immer gleich auszuführen, was die Quantität und Qualität signifikant steigert. Sich bloss auf Stichproben zu beschränken, wird mit RPA überflüssig. Zudem wird in der Regel eine deutlich stärkere Standardisierung in den Arbeitsprozessen erreicht, was dem Unternehmen etwa bei Personalwechsel sehr zugutekommt.

10.2 Aufbau

Für den Aufbau von RPA müssen Sie einen Prozess in seine kleinsten Teilschritte aufgliedern. Definieren Sie für jeden Teilschritt die Bedingungen im Sinne von:

- Wie wird der Teilschritt gestartet?
- Wann wird der Teilschritt gestartet?
- Ist der Teilschritt abhängig von einem vorgängigen Schritt?
- Kann der Teilschritt mit einem anderen Teilschritt parallel durchgeführt werden?
- Was passiert, wenn der Teilschritt abbricht oder andere technische Probleme verursacht?
- Wie ist der Teilschritt zu protokollieren?
- Braucht der Teilschritt eine manuelle Interaktion?

Die meisten RPA-Tools stellen die Definition des Prozesses in einem Ablaufdiagramm dar, so sind die Bedingungen auf einem Blick erkennbar. Spätestens in der Definition eines Teilschritts wird ersichtlich, dass sich die Automatisierung von Prozessen nur bei hoher Standardisierung empfiehlt.

10.3 Robotic Process Automation bei einer Kundendatenbank

Die Einsatzmöglichkeiten von RPA bei einer Kundendatenbank sind vielfältig und daher individuell auf das Unternehmen zuzuschneiden. Um Ihnen ein paar Ideen mitzugeben, werden im Folgenden einige mögliche Anwendungsfälle im Zusammenhang mit einer Kundendatenbank beschrieben. Die Anwendungsfälle sind nach Komplexität der Kundendatenbank sortiert, sprich von der einfachen Kundenliste zur State-of-the-Art-Kundendatenbank und BI-Lösung.

Integritätskontrolle

In einer einfachen Kundenliste ist es essenziell, dass die Liste in sich integer ist. Datenintegrität heisst, dass die Kundendaten in sich valide sind und sich nicht widersprechen. Mit RPA kann beispielsweise vollautomatisch geprüft werden, ob ein Kunde mehrfach in der Kundenliste hinterlegt und ob jedes Muss-Feld (Adressen, Kunden-Nr etc.) ausgefüllt ist. Offene Rechnungsbeträge können mit dem Stand der Forderungen aus Lieferungen und Leistungen abgeglichen werden. Strassen und weitere Adresselemente können mit einem externen Adressverzeichnis geprüft werden. Festgestellte Abweichungen können einem User zur Datenbereinigung disponiert oder bis zu einem gewissen Grad direkt vom System eigenständig korrigiert werden.

Übernahme von Daten

Es kann vorkommen, dass Kundendaten manuell in der Kundendatenbank erfasst werden müssen (z.B. ein Kunde meldet sich über E-Mail). RPA kann dabei unterstützen, indem sämtliche Kunden-Mails vollautomatisiert in strukturierte Daten umgewandelt und in die Datenbank geschrieben werden. Es wäre auch denkbar, dass sich die Datenerfassung von telefonischen Kundenanmeldungen anhand von Stimmerkennung und weiteren Daten (z.B. anrufende Telefonnummer) mit RPA automatisieren lässt.

Konsolidierung

Bei einer Fusion zweier Unternehmen müssen die Kundenbestände zusammengeführt und konsolidiert werden. Ein aufwendiges Unterfangen, das unter Umständen eine lange Durchlaufzeit beansprucht und das operative Tagesgeschäft behindern kann. Mit RPA können komplexe Migrationslogiken automatisiert werden – damit lässt sich die Durchlaufzeit um ein Vielfaches reduzieren, was schlussendlich auch die Qualität deutlich steigert.

Kundenaktionen

Es besteht in der Regel das Bedürfnis, regelmässig mit den Kunden in Kontakt zu treten, etwa durch Newsletter über E-Mail und Ähnliches. Mit RPA können Massenverarbeitungen wie E-Mails, Briefe, Online-Posts und weitere Kommunikationen vollautomatisiert wiederkehrend durchgeführt werden.

Automatisierungen von BI-Lösungen

Wenn die Kundendatenbank für Auswertungen mit externen Datenquellen verknüpft und umfangreiche BI-Lösungen aufgebaut werden, wird oft ein Data Warehouse eingesetzt, um eine performante Auswertung sicherstellen zu können. Ein Data Warehouse erlaubt auch, eine operative Kundendatenbank strukturell für Auswertungen aufzubereiten und eine stärkere Historisierung zu ermöglichen. Beispielsweise soll eine Kundenbestellung mit der im Bestellzeitpunkt gültigen Adresse des Kunden verknüpft werden.

In der operativen Datenbank liegt der Fokus auf dem Tagesgeschäft und die Datenbank soll nicht durch umfangreiche Relationen aufgebläht werden. Der Einsatz eines Data Warehouse erfordert eine Vielzahl an technischen Jobs, die die Daten extrahieren, transformieren und in das Data Warehouse laden (ETL). Zudem besteht oft das Bedürfnis, nach der Datenaufbereitung die Datenkonsumenten auf die neuen Auswertungen hinzuweisen oder sie gleich mit entsprechenden Dateien zu bedienen. Es ist auch vorstellbar, gleich ein Dashboard für das Management aufzubauen, das periodisch mit aktuellen Kennzahlen geladen wird.

In dieser ganzen Prozesskette der Datenanreicherung, -auswertung und -publikation kann RPA die vielen einzelnen Jobschritte vollautomatisch orchestrieren. Auch Toleranzwerte können definiert werden, und wird ein Toleranzwert verletzt, so meldet der Roboter automatisch einen Hinweis an das Controlling.

10.4 Robotic Process Automation bei Softwaretests

Je individueller und komplexer Datenbanken und Datenauswertungen sind, desto höher ist der Testaufwand bei neuen Releases und ähnlichen Weiterentwicklungen. Da Kundendatenbanken zu den sehr geschäftskritischen Bereichen zählen, sind entsprechende Veränderungen einer vertieften Qualitätssicherung zu unterstellen. Mit RPA können Usereingaben x-fach durchgeführt und systematisch nach Fehlern ausgewertet werden. Wenn solche Tests im Deployment-Prozess eingebunden sind, kann vollautomatisch sichergestellt werden, dass jede Softwareveränderung eine Reihe von Tests durchlaufen muss, bevor diese produktiv aufgeschaltet wird.

10.5 Einrichtung von Robotic Process Automation – was gilt es zu beachten?

RPA einzuführen, erfordert eine projektorientierte Vorgehensweise. Denken in Prozessschritten ist essenziell – bitte beachten Sie auch folgenden Punkte:

1. Wählen Sie ein Standardprodukt

Zu RPA gibt es auf dem Markt viele Standardprodukte, vermeiden Sie, wenn immer möglich, etwas selbst entwickeln zu wollen.

2. Beispiele von RPA-Anbietern

- UiPath (Gartner, 2020)
- Automation Anywhere (Gartner, 2020)
- Automate Software

3. Sehr tiefe Payback-Dauer

Bevor Sie in RPA investieren, erstellen Sie eine Investitionsrechnung. RPA kann rasch hohe Investitionen erfordern, die wirtschaftlich sein sollten. Software ist kurzlebig, eine Amortisationsdauer von mehr als zwei bis drei Jahren ist kritisch. Sollen auch nicht monetäre Gründe (z. B. Know-how-Sicherung) einbezogen werden, so empfiehlt sich zusätzlich eine Nutzwertanalyse.

4. Etappiertes Vorgehen

Versuchen Sie nicht, sämtliche Prozesse oder ganze Prozessketten auf einmal zu automatisieren. Gliedern Sie einen Prozess in Teilprozesse, die Sie dann etappiert automatisieren, um rasch Erfahrungen zu sammeln und darauf den Umfang der angestrebten RPA-Lösung justieren zu können.

5. Bewusst manuelle Schritte

RPA darf nicht dazu führen, dass Sie die Kontrolle über die Prozesse verlieren. Bauen Sie bewusst manuelle Schritte ein, um IKS-Anforderungen wie etwa das 4-Augen-Prinzip einhalten zu können (z. B. Freigabe einer Kundengutschrift).

6. Berechtigung inkl. Protokollierung

Die RPA-Software benötigt oft weitreichende Berechtigungen in verschiedenen Systemen. Daher ist es notwendig, ein stringentes Berechtigungskonzept zu erstellen, um den Menschen hinter der RPA-Software klar zu identifizieren. Sämtliche Jobschritte sollten auch protokolliert werden, damit Sie im Nachhinein die Transaktionen nachvollziehen können.

7. Error-Handling

RPA-Prozesse laufen oft ausserhalb von Geschäftszeiten, um das operative Geschäft nicht negativ zu beeinflussen (etwa durch erhöhte Systembelastungen). Der Reaktionsfähigkeit in einem Fehlerfall soll daher besondere Beachtung geschenkt werden. Was muss bei einem Job-Abbruch getan werden? SMS-Mitteilungen u. ä. können bei jedem Job-Schritt für den Fehlerfall hinterlegt werden. Reaktionszeiten ausserhalb von Geschäftszeiten sind gesondert zu vereinbaren.

- 8. Seien Sie sich auch bewusst:** RPA verträgt keine Abweichung von der Norm. Findet die Software nicht die erwartete Ausgangslage vor, so bricht der Teilschritt ab. Definieren Sie in diesem Fall die Massnahmen wie etwa E-Mail-Benachrichtigungen an den Administrator o. ä.

9. Durchlaufzeit

Achten Sie bei RPA auf die Durchlaufzeit. Je länger ein automatisierter Prozess an Durchlaufzeit benötigt, desto höher ist das Risiko für technische Probleme. Erstens steigt mit der Durchlaufzeit die Häufigkeit von Timeouts und ähnlichen unerwünschten Situationen. Zweitens wird der negative Einfluss auf das operative Tagesgeschäft deutlich frappierender, je länger die Durchlaufzeit und damit die Wiederherstellung dauert. Es ist empfehlenswert, von Beginn an Zielgrössen für die Durchlaufzeit festzulegen und die effektiven Durchlaufzeiten regelmässig mit dem Zielwert zu messen.

10. RPA ist meistens nicht nur ein Tool

RPA-Lösungen bestehen oft aus verschiedenen Tools. Meist wird eine Software eingesetzt, die mit ganz verschiedenen Applikationen kommunizieren kann und die einzelnen Schritte der Applikationen terminiert in einen Ablauf bringt. Zusätzlich werden dann in den Teilschritten spezifische Tools eingesetzt, etwa ein Tool, das sich auf Vergleiche und Abgleiche spezialisiert hat. Erst zusammen kann erreicht werden, Abgleiche von Neben- und Hauptbüchern zu automatisieren.

11. Weniger Doing, aber mehr Monitoring

RPA entlastet das Controlling in der Routinearbeit, erhöht aber den Überwachungsaufwand. Versuchen Sie, den Kontrollaufwand möglichst effizient zu halten. So kann etwa dem Controlling auch ein Monitoring-Cockpit zur Verfügung gestellt werden, das auf einen Blick den Status aller RPA-Prozesse anzeigt.

12. Die Grenzen kennen

Die Interpretation und Kommentierung von Daten lassen sich schwer vollständig automatisieren – da braucht es das Controlling und das auf das Geschäft zugeschnittene Know-how.

11 Künstliche Intelligenz und Business Analytics

11.1 Was ist künstliche Intelligenz?

Wenn wir heute in eine Suchmaschine «Konto» eingeben, wird uns eine Vervollständigung wie «Konto eröffnen» vorgeschlagen. Wenn wir heute ein Bild aus dem Urlaub auf einen Cloud-Foto-Dienst laden, wird als Stichwort «Strand» vorgeschlagen. Wenn wir uns heute für eine Kreditkarte bewerben und unsere Daten eingeben, wird für uns eine Kreditwürdigkeit berechnet. Wenn wir uns heute für ein Abo jeglicher Art einschreiben, erhalten wir regelmässig vom Anbieter des Abos auf uns zugeschnittene Empfehlungen.

Alle diese Vorgänge haben eine Gemeinsamkeit: Den Modellen, die diese Vorschläge und Berechnungen durchführen, wurden Daten gegeben, um Muster zu erkennen. Die Suchmaschine lernt auf Basis historischer Suchvorgänge, dass viele Nutzer nach «Konto» «eröffnen» schreiben. Der Cloud-Foto-Dienst hat nach einer Vielzahl an Fotos, die mit dem Tag «Strand» versehen worden sind, gelernt, wie ein Strand aussieht, und kann jetzt für jedes Foto einschätzen, ob ein Strand zu sehen ist oder nicht. Ein Modell zur Bewertung der Kreditwürdigkeit hat auf Basis von zuvor bestimmten Kreditwürdigkeiten für Personen gelernt, dass einer Person mit hohem, stabilem Einkommen eine höhere Kreditwürdigkeit gegeben wird als einer Person, die gerade aus der Lehre gekommen ist. Abonnenten werden auf Basis ihrer persönlichen Daten und Aktivitäten in Gruppen oder Segmente eingeteilt, da Anbieter die Daten von Tausenden von Abonnenten besitzen und somit Muster innerhalb der Abonnenten erkennen konnten.

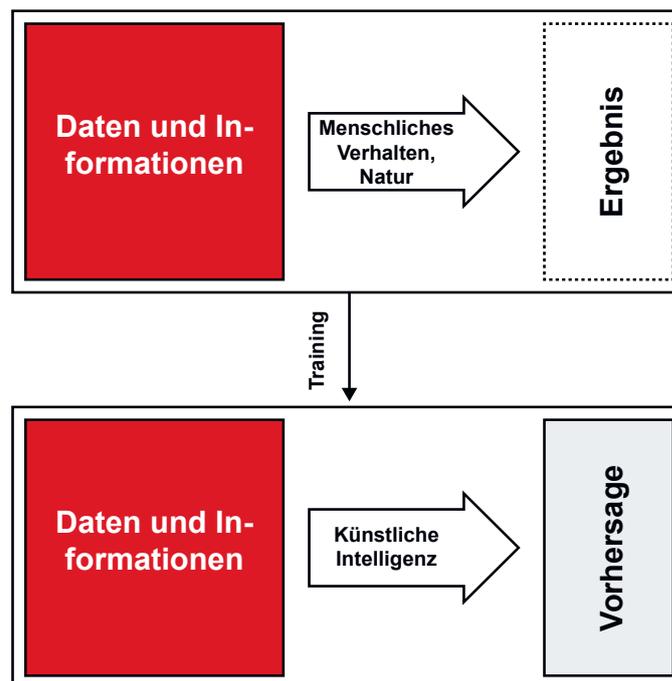


Abbildung 13: Natur einer Künstlichen Intelligenz

Anhand von historischen Daten und Informationen werden Verhaltensmuster, Naturgesetze oder andere Regeln imitiert.

Künstliche Intelligenz, kurz KI (engl. AI) beschreibt den Vorgang, Muster zu erkennen und anschliessend auf neue Daten anzuwenden, um mithilfe der erkannten Muster eine **Vorhersage** oder Einschätzung zu machen. Die Mustererkennung wird auch als **Lernen** oder **Training** bezeichnet, ein KI-Modell wird also **trainiert**. Da das Training zumeist nicht per Hand, sondern durch einen Computer stattfindet, spricht man auch von

maschinellern Lernen oder **Machine Learning**. Im täglichen Sprachgebrauch werden ausserdem Bezeichnungen wie Advanced Analytics, Data Mining und Predictive Analytics synonym verwendet, obwohl insbesondere die letzten drei Bezeichnungen eigentlich Teilbereiche der KI darstellen.

Um diese Mustererkennung noch besser zu verstehen, generieren wir uns ein künstliches Beispiel. Nehmen wir an, wir vermarkten einen Kredit und wollen diesen Kredit nur an Personen weitergeben, bei denen wir wissen, dass sie ihn zurückzahlen können. Wir verfügen lediglich über drei Informationen für jeden Bewerber, nämlich ob die Person bereits ein Konto bei uns besitzt, ob sie die Farbe Orange mag und wie oft die Person im Jahr spazieren geht. Der folgende historische Datensatz liegt uns vor:

Bereits Kunde?	Mag Orange?	Spaziergänge pro Jahr?	Kredit zurückgezahlt?
Ja	Nein	3	Ja
Nein	Nein	9	Nein
Nein	Nein	10	Ja
Ja	Ja	5	Ja
Nein	Ja	9	Nein
Ja	Nein	15	Ja
Ja	Ja	10	Ja
Nein	Ja	11	Nein
Nein	Nein	8	Nein
Nein	Nein	11	Ja

Versuchen wir, Muster zwischen den drei Informationen und dem Ergebnis *Kredit zurückgezahlt* zu erkennen, so werden wir schnell fündig. Ist der Applikant bereits Kunde, dann wird der Kredit immer zurückgezahlt. Ist der Applikant noch nicht Kunde und mag die Farbe Orange, dann wird der Kredit nie zurückgezahlt. Ist der Applikant noch nicht Kunde und mag Orange nicht, dann wird der Kredit zurückgezahlt, wenn er mindestens 10-mal pro Jahr einen Spaziergang macht. Dieses (zugegebenermassen etwas realitätsfremde) Beispiel zeigt anschaulich, was es für uns bedeutet, Muster zu erkennen. Genau das macht eine KI auch. Die KI versucht, Regeln zu erkennen, nach denen sie die uns vorliegenden Informationen verarbeiten und anschliessend eine Vorhersage generieren kann. Wollen wir also wissen, ob ein uns noch unbekannter Applikant, der Orange nicht mag und 12 Spaziergänge pro Jahr macht, den Kredit zurückzahlen kann, so ist die KI nun, nachdem sie die Muster in einem Training erkannt hat, in der Lage, unsere Gemüter zu beruhigen und vorherzusagen, dass diese Person den Kredit vermutlich zurückzahlen wird.

Natürlich kann KI auch komplexere Situationen erkennen, die wir – wenn überhaupt – nur mit sehr viel Aufwand erkennen könnten. In genau diesen Situationen vereinfacht KI uns das Leben. Damit wir für die kommenden Herausforderungen als Accountant im Controlling oder anderen Domänen ein besseres Verständnis der vielen verschiedenen Facetten einer KI haben und somit besser beurteilen können, wo es Sinn macht, KI einzusetzen und was dabei zu beachten ist, möchten wir uns hier gerne mit diesem Thema auseinandersetzen.

In diesem Kapitel analysieren wir die verschiedenen Facetten von KI. Im Abschnitt «Die verschiedenen Formen der KI» bringen wir Struktur in die weite Welt der KI und führen wichtige Begriffe ein. Anschliessend diskutieren wir im Abschnitt «Die Entstehung von KI-Modellen», wie KI-Modelle entstehen, was eine gute KI ausmacht und wie wir sie evaluieren und validieren können. Risiken von KI-Modellen werden in «Die Anfälligkeiten eines KI-Systems» umrandet. Schlussendlich wird in «Der schnelle Weg zum ersten Modell» beleuchtet, mit welchen Werkzeugen wir heutzutage in kürzester Zeit zu einem ersten (vielleicht noch nicht ganz perfekten) Modell kommen können.

Wichtige Begriffe aus diesem Abschnitt: Lernen, Training, Vorhersage, KI, AI, Maschinelles Lernen, Machine Learning, Advanced Analytics, Data Mining, Predictive Analytics.

11.2 Die verschiedenen Formen der KI

Die oben aufgeführten Beispiele umrahmen die Kernidee von KI: das Erkennen von Mustern und das Anwenden dieser erkannten Muster auf neue Fälle. Dennoch unterscheiden sich die einzelnen Fälle in vielfacher Hinsicht und geben einen kleinen Einblick in die Vielfältigkeit von KI.

Überwachtes und unüberwachtes Lernen

Eine grundlegende Unterscheidung von KI liegt darin, worauf sie lernen. Diese Daten bezeichnen wir als **Trainingsdaten**. Trainingsdaten enthalten in jedem Fall die Daten, in denen Muster erkannt werden sollen. Für das Beispiel der Kreditwürdigkeit kann dies eine Vielzahl an Daten umfassen: Einkommen, Vermögen, Dauer Sesshaftigkeit in aktueller Gemeinde, Anzahl Jobs in den letzten fünf Jahren und vieles mehr. Wir bezeichnen dieses Daten, die als Eingabe in das Modell gehen, als **Features**.

Das Beispiel der Kreditwürdigkeit enthält aber mehr als nur Features. Die Trainingsdaten umfassen auch die Kreditwürdigkeit, oftmals in Form von CHF. Dieses «Ergebnis» haben wir nur für Trainingsdaten – bewirbt sich eine neue Kundin um ein Konto, so ist ihre Kreditwürdigkeit noch nicht bekannt. Dieses «Ergebnis» bezeichnen wir auch als **Label**. KI, die auf Trainingsdaten trainiert wird, die neben Features auch Labels besitzen, wird auch als **überwachtes Lernen** bezeichnet. «Überwacht» bezieht sich dabei einfach auf die Begebenheit, dass ein Ergebnis, das Label, bereits vorhanden ist. Das Erkennen von Objekten auf Bildern fällt auch in die Kategorie des überwachten Lernens, schliesslich lernt das Modell basierend auf Bildern, die mit Stichworten (Labels) wie Strand, Palme, Berge etc. versehen sind.

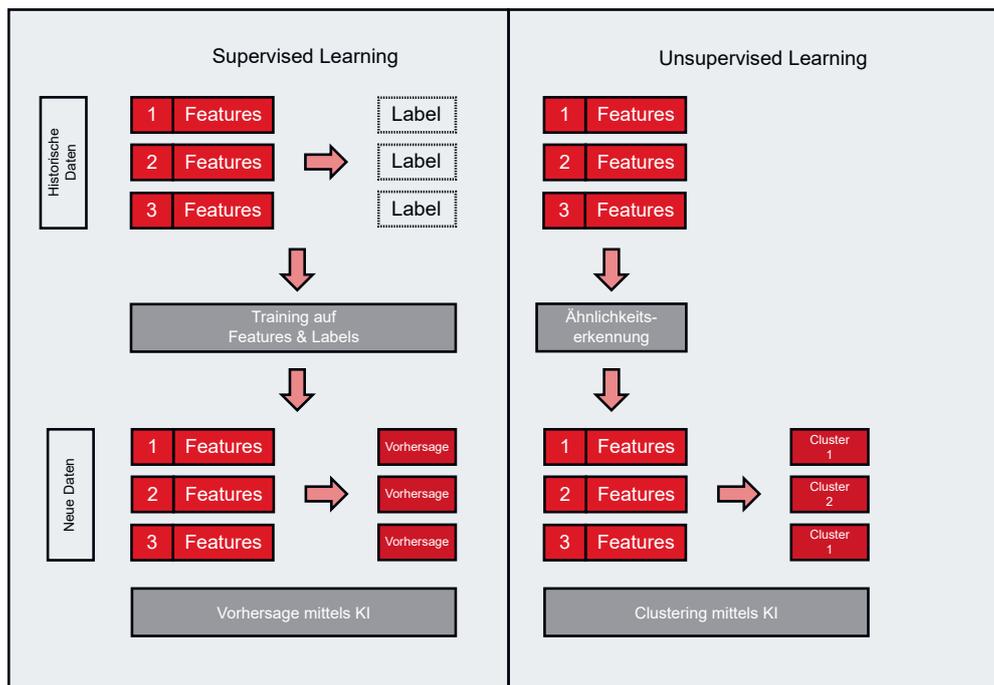


Abbildung 14: Differenz Supervised Learning/Unsupervised Learning

Die Illustration bildet die grundlegende Differenz zwischen Supervised Learning (links) und Unsupervised Learning ab. Ersteres besitzt für jeden Datensatz (z. B. jeden Interessenten für ein Konto) a priori ein Label (z. B. Ist/Nicht kreditwürdig) – mittels KI gilt es die Regeln, die das Mapping von Features zu Label bestimmt haben, zu imitieren, sodass auch wir eine Vorhersage zur Kreditwürdigkeit machen können. Letzteres besitzt in den Daten, auf denen trainiert wird, kein Label, es existiert also a priori kein Ergebnis. KI versucht, ein Label aus den Daten a posteriori zu implizieren, etwa die Zugehörigkeit zu einem Cluster, die erst durch KI ersichtlich wird.

In einigen Beispielen existieren solche Labels a priori nicht. Nehmen wir also an, wir sind ein Mobilfunkanbieter und wollen unsere Kunden besser verstehen. Natürlich macht es keinen Sinn, jeden einzelnen Kunden im Detail zu betrachten und zu verstehen – das könnte etwas länger dauern. Stattdessen wollen wir unsere Kunden in Gruppen aufteilen, die jeweils mehr oder weniger homogen sein sollen. In anderen Worten: Wir suchen Cluster, in denen die Kunden viele Ähnlichkeiten miteinander teilen. Diese Art von KI, die Muster/Gruppen/Cluster/Segmente erkennt, wird auch als **unüberwachtes Lernen** bezeichnet, da es kein Label gibt. Auch im Beispiel neuer Interessenten für ein Konto kann unüberwachtes Lernen angewendet werden. Die Trainingsdaten könnten die oben genannten Features umfassen, und das KI-Modell könnte darin Gruppierungen erkennen, wie etwa die *Langansässigen mit stabilem Einkommen und moderatem Vermögen*, die *Selbstständigen mit instabilem Einkommen und hohem Vermögen* oder auch die *Rentner mit wenig, aber stabilem Einkommen und langer Ansässigkeit*.

Unüberwachtes Lernen oder *Unsupervised Learning* umfasst aber mehr als nur die Erkennung von Clustern. Die Erkennung von Ausreißern, die sogenannten *Outlier* oder *Anomaly Detection* sowie *Association Rules* gehören auch zum Teilgebiet des *Unsupervised Learning*, da die zugrunde liegenden Trainingsdaten a priori nicht mit einem Label versehen waren.

Wichtige Begriffe aus diesem Abschnitt: Trainingsdaten, Features, Label, überwachtes Lernen, supervised learning, unüberwachtes Lernen, unsupervised learning, Outlier detection, anomaly detection, association rules.

11.3 Klassifizierung und Regression

Eine weitere Unterscheidung der verschiedenen Formen der KI kann im Beispiel des überwachten Lernens auch mit Bezug auf das Label gemacht werden. Nehmen wir uns wieder des Beispiels der Kreditwürdigkeit an. Wie zuvor beschrieben, können die Features etwa das Einkommen, das Vermögen oder andere persönliche, finanzielle Daten umfassen. Das Label kann z. B. durch die Kreditwürdigkeit in CHF beschrieben werden. Wir betrachten also ein Label, das den kontinuierlichen Bereich zwischen einem Minimalwert (z. B. 0 CHF) und einem Maximalwert (z. B. CHF 1'000'000) beschreibt. In einem alternativen Szenario könnte es allerdings reichen, wenn man die Frage der Kreditwürdigkeit mit einem «Ja» oder «Nein» beantwortet. Die Trainingsdaten würden dann die gleichen Features wie oben beschreiben umfassen, das Label würde dann aber nur eines der beiden Werte «Ja» oder «Nein» umfassen. Während wir das erste Beispiel einer KI, die die Kreditwürdigkeit zwischen CHF 0 und CHF 1'000'000 vorhersagt, als **Regression** bezeichnen, verwenden wir die Bezeichnung **Klassifizierung** oder **Classification** für Fälle, in denen aus zwei Klassen (hier: «Ja» oder «Nein») die richtige vorhergesagt werden soll. Multilabel Classification beschreibt die Klassifizierung mit mehr als zwei Klassen, also etwa eine Bilderkennungs-KI, die schätzt, ob auf einem Bild eine Palme, ein Strand ein Berg und/oder eine Grotte zu sehen ist.

Wichtige Begriffe aus diesem Abschnitt: Regression, Klassifizierung, Classification.

Tabellarisches Maschinelles Lernen, Bilderkennung und Texterkennung

KI kann auch nach Anwendungsfällen kategorisiert werden. Die drei grössten Gebiete umfassen **Texterkennung (bzw. Spracherkennung)** oder **Natural Language Processing (NLP)**, **Bilderkennung** oder **Visual Recognition** und **Tabellarisches Maschinelles Lernen** oder **Tabular Machine Learning**. NLP und Visual Recognition simulieren dabei menschliche Kognitionsfähigkeiten, was einem «klassischen Verständnis» von KI entspricht.

Tabellarisches Machine Learning umfasst alle Beispiele, in denen die Features typischerweise tabellarisch dargestellt werden können. Der Fall der Kreditwürdigkeit ist ein klassisches Beispiel: Ein Datensatz entspricht einer Reihe, dessen Spalten die Features, also z. B. Einkommen, Vermögen etc. und das Label darstellen – eine tabellarische Ausführung.

Bildererkennung beschreibt das Lernen basierend auf einer Zahl an Bildern mit Label. Das Label eines Bildes kann dabei festhalten, welches Objekt auf dem Bild zu sehen ist. Alternativ kann das Label auch den Ort eines Objekts auf dem Bild umfassen, um eine Objekterkennung zu ermöglichen. In ersterem Fall wird das trainierte Modell dazu verwendet, um Bilder, die noch kein Label umfassen, mit einer Einschätzung anzureichern. Eine App, die beim Fotografieren einer Blume den Namen dieser Blume wiedergibt, basiert genau auf diesem Konzept. Die App wurde zuvor auf einem grossen Datensatz an Bildern von Blumen mit einem Label – etwa Gänseblümchen oder Pustebblume – für jedes Bild trainiert.

Texterkennung beschreibt die Erkennung von Mustern aus Text. Dies kann verschiedene Formen annehmen. Ein klassisches Beispiel umfasst die Übersetzung von Texten. Bekannte Online-Anbieter wie DeepL basieren dabei auf einem eigentlich recht einfachen Prinzip: Die Trainingsdaten bestehen aus Texten von Websites, die sowohl auf einer als auch auf einer anderen Sprache geschrieben worden sind. Auf dieser Basis kann die Texterkennung dann lernen, wie Wörter und Sätze zu übersetzen sind. Ein anderes, klassisches Anwendungsgebiet ist die sogenannte Sentiment Analysis. Hier wird ein Satz auf positive oder negative Orientierung hin überprüft, was etwa im Support-Service-Team eines Unternehmens nützlich sein kann, da somit auf Nachrichten von Nutzern bessere automatisierte Antworten geliefert werden können.

Wichtige Begriffe aus diesem Abschnitt: Texterkennung, Spracherkennung, Natural Language Processing (NLP), Bildererkennung, Visual Recognition, Tabellarisches Maschinelles Lernen, Tabular Machine Learning.

Künstliche Intelligenz als Basis für Automatisierung/Unterstützung

Wir wollen noch eine letzte Differenzierung diskutieren. Künstliche Intelligenz wird häufig als Basis für *Automatisierung* gesehen. Das ist auch teilweise korrekt, etwa wenn ein Texteditor auf Basis historischer Daten automatisch Korrekturen vornimmt oder ein Kundensystem bei hoher Abgangswahrscheinlichkeit eines Kunden (berechnet via KI) automatisch eine Mail entsendet. Künstliche Intelligenz muss aber nicht unbedingt Prozesse ganzheitlich ablösen, in vielen Fällen macht es mehr Sinn, das Ergebnis des KI-Modells als *Unterstützung* für einen Entscheidungsprozess zu nehmen. Auch hier hilft uns das Beispiel der Kreditwürdigkeit. Ein KI-Modell könnte für jeden Interessenten eine Kreditwürdigkeit in CHF berechnen und dann automatisch einen Vertrag generieren. Alternativ könnte das Ergebnis aber auch einem Bankangestellten als Basis für seine Überlegungen und Vertragsverhandlungen mit dem Kunden dienen – wie er diese Zusatzinformation benutzt, ist letztendlich immer noch ihm überlassen.

11.4 Die Entstehung von KI-Modellen

Wir haben bisher verschiedene Beispiele gesehen, bei denen ein Modell auf Basis historischer Daten Muster erkennt und später, auf neuen Daten, ebendiese Muster anwendet, um Vorhersagen/Einschätzungen treffen zu können. Innerhalb dieses Prozesses müssen jedoch verschiedene Bereiche betrachtet werden:

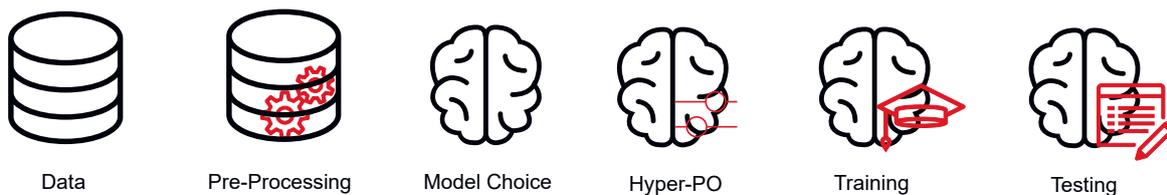


Abbildung 15: Schritte von Daten bis zum Testing

- 1. Data Pre-Processing:** Wie können wir unsere Daten anpassen, sodass das Modell bestmöglich Muster erkennen kann?

2. **Training und Modellauswahl:** Welche Daten geben wir dem Training? Welches Modell passt am besten zum Problem, das wir behandeln? Kann ich das Modell konfigurieren und wenn ja, was ist die beste Konfiguration?
3. **Testing:** Welche Daten nehme ich für das Testing? Wie messe ich die Güte des Modells oder anders formuliert, wann ist das Modell ein gutes Modell?

In diesem Abschnitt möchten wir auf diese drei Bereiche eingehen, die wir auf dem Weg zu unserem finalen Modell adressieren müssen.

Data Pre-Processing

Daten stellen die Basis dar, mit dem wir ein Modell füttern, sodass dieses Muster daraus erkennen kann. Es gilt das sehr einfache Prinzip *Garbage in, garbage out*. Solange Daten die Muster nicht zu erkennen geben oder nicht in einer Form sind, in der das Modell diese Muster entnehmen kann, wird das Modell diese Muster auch nicht erkennen.

Es liegt also an uns, die Daten so herzurichten, dass das Modell Muster entnehmen kann. Dabei können wir aus einer Vielzahl an Werkzeugen auswählen. Wir gehen hier auf die wichtigsten Methoden ein:

- **Filterung von Datensätzen** oder *wie erhalte ich einen Schönwetterdatensatz?* Eine Filterung der Daten erlaubt es, nur die wichtigsten und qualitativ hochwertigsten Daten beizubehalten. Das kann insofern helfen, als dass in diesen hochwertigen Daten Muster einfacher zu erkennen sind. Insbesondere die sogenannten Aussenseiter oder Outliers wollen wir damit beseitigen. Dies bezeichnen wir auch als Anomaly Detection – ein Schritt, für den es zahlreiche vorgefertigte, öffentlich zugängliche Bibliotheken gibt. Bei der Filterung von Daten ist zu beachten, dass dies im Rahmen des Trainings geschieht. Im Betrieb kann diese Filterung entweder übernommen und mit gewissen Regeln versehen werden oder auch nicht. In letzterem Fall müssen wir jedoch analysieren, wie das Modell auf diese Anomalien reagiert – schliesslich kennt es sie nicht aus dem Training.
- **Anreicherung von Datensätzen** oder *wie kann ich mit fehlenden Daten umgehen?* Manchmal fehlt es an Daten. Betrachten wir beispielsweise die Erkennung von *fraudulent transactions* – also Transaktionen illegaler Natur –, dann kommen diese sehr viel seltener vor als herkömmliche Transaktionen. Für ein Modell kann es jedoch zu Schwierigkeiten führen, wenn es von einer Klasse (*fraudulent transactions*) sehr viel mehr gibt als von der anderen. In diesem Fall bieten sich mehrere Optionen an. Zwei von diesen Optionen umfassen die Reduktion der Testdaten der grösseren Klasse oder der Replikation von Testdaten der kleineren Klasse. Diesen Vorgang nennen wir Resampling. Natürlich gibt es noch zahlreiche weitere Verfahren wie etwa das Generieren synthetischer Datensätze mithilfe von Techniken wie SMOTE (*Synthetic Minority Oversampling Technique*), aber für den Umfang dieses Skripts belassen wir es hierbei.
- **Feature Engineering** oder *wie können wir unsere Spalten so übersetzen und erweitern, dass auch das Modell sie versteht?* Oftmals erhalten wir in unseren Datensätzen Spalten (oder *Features*), die Daten in einem Format enthalten, das für das Modell nicht leicht zu interpretieren ist. Nehmen wir beispielhaft an, bei der Anmeldung eines Kunden für ein Konto gibt dieser in einem Freitextfeld all seine bevorzugten Anlagentypen an, z. B. «Aktien, Immobilien, Kryptowährung», ein anderer gibt an «Aktien und Immobilien» und wieder eine andere Person gibt an «Aktien und Kryptowährung». Für das Modell sind diese Einträge nichts anderes als verschiedene Klassen. Um dem Modell beim Verstehen zu helfen, dass es sich hier um einige wenige Klassen («Aktien», «Immobilien», «Kryptowährung») handelt, die sich immer wieder wiederholen, müssten wir neue Features erstellen wie «Hat bevorzugten Anlagentyp Aktien», «Hat bevorzugten Anlagentyp Immobilien», «Hat bevorzugten Anlagentyp Kryptowährung». Für uns tragen diese neuen Features keine neuen Informationen. Für das Modell erlauben diese neuen Features allerdings, Muster einfacher zu erkennen, die mit dem bevorzugten Anlagentyp zu tun haben.

- **Imputing** oder *was machen wir mit leeren Feldern?* Imputing bezeichnet das Füllen fehlender Daten. Nehmen wir beispielsweise an, das Feld *Alter* im Anmeldeformular für ein Konto ist nicht obligatorisch. Die Mehrzahl der Kunden füllt es nun aus, die anderen nicht. Dem Modell können jedoch keine leeren Datenfelder gegeben werden bzw. wir müssen dem Modell mitteilen, wie es diese zu verstehen hat. Wir könnten z. B. versuchen, das Feld durch einen konstanten Wert, etwa den Durchschnitt der Trainingsdaten mit Werten, zu ersetzen. Dies ist jedoch nicht immer zu empfehlen bzw. der konstante Wert ist mit Vorsicht auszuwählen. Alternativ könnten wir basierend auf den restlichen Werten, die im Anmeldeformular angegeben worden sind, das Alter schätzen. Dies ergibt insofern eine interessante Konstellation, als dass wir KI verwenden, um Daten zu bereinigen, die für das Trainieren einer anderen KI genutzt werden.
- **Normalization** oder *warum auch das Modell Ähnlichkeit schätzt*. Normalization bezeichnet die Normalisierung von Daten, also insbesondere das Stauchen und Strecken der Einträge in den verschiedenen Spalten, sodass z. B. der Mittelwert und die Standardabweichung der Spalten mit kontinuierlichen Werten übereinstimmt. Warum ist das aber wichtig? Bedienen wir uns ein weiteres Mal unseres Beispiels rund um die Kreditwürdigkeit eines möglichen Kunden. Ein Feature dieses Kunden könnte in seinem Alter bestehen. Ein Wert, der etwa zwischen 18 und 100 variiert. Ein weiterer Wert kann im Vermögen eines Kunden bestehen. Dieser kann 0 sein, für viele wird er im Bereich von CHF 10'000 bis CHF 1'000'000 liegen, es wird aber sicherlich auch Kunden mit einem Vermögen von mehr als CHF 1'000'00 geben. Wiederum eine andere Spalte kann die durchschnittliche Note in der letzten Schulklasse eines Kunden darstellen. Diese Note liegt zumeist zwischen 4 und 6. Verschiedene KI-Modelle können mit diesen unterschiedlichen Intervallen unterschiedlich gut umgehen. Für die einen bedeuten sie lediglich eine längere Trainingszeit, um Muster zu erkennen, andere schaffen dies kaum oder nur mit ungenügender Qualität. In der Regel bietet es sich deshalb an, die jeweiligen Spalten so zu skalieren und zu verschieben, dass der Mittelwert bei 0 und die Standardabweichung bei 1 liegt. Dies erlaubt, eine gewisse Ähnlichkeit in der Verteilung der Daten zu forcieren – das Modell nimmt dies zumeist dankend in Form von kürzerer Trainingszeit und besseren Resultaten an.

Wichtige Begriffe: Pre-Processing, Aussenseiter, Outlier, Anomaly Detection, Datenanreicherung, Oversampling, Feature Engineering, Imputing, Normalization.

Training und Modellauswahl

Sobald die Daten bereit sind, kann das Modell trainiert werden. Zuvor muss jedoch noch das passende Modell ausgewählt werden. Für die verschiedenen Arten der künstlichen Intelligenz, gibt es verschiedene Modelle. Hier eine kleine Auswahl der Modelle, die man kennen sollte:

- **Klassifizierung:** Decision Tree, Random Forest, Xgboost, Logistic Regression, GLM, Neural Networks
- **Regression:** Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression, Bayesian Linear Regression, Elastic Net Regression
- **Clustering:** KMeans, Nearest Neighbor

Die einzelnen Modelle zu verstehen, würde den Rahmen dieses Skripts sprengen. Wir fokussieren uns hier auf ein einziges Modell, um zu verstehen, was Künstliche Intelligenz bzw. die Erkennung von Mustern eigentlich bedeutet: den Decision Tree.

Ein Decision Tree (dt. Entscheidungsbaum) kann als eine Reihe von Fragen an die Daten verstanden werden, von der jede mit Ja oder Nein beantwortet wird. Je nachdem welche Sequenz an Antwort gegeben wird, resultiert eine Klasse. Abbildung X illustriert dies anhand des Beispiels der Kreditwürdigkeit. Nehmen wir an, wir sind Student mit einem jährlichen Einkommen von CHF 30'000, und wir haben keine Schulden. Der erste Knoten (engl. node) fragt, ob das Einkommen höher als CHF 80'000 liegt. Da dies nicht der Fall ist, gehen wir zum rechten Knoten, der fragt, ob wir Student sind. Dies beantworten wir mit Ja, weshalb eine weitere Frage aufkommt, nämlich die der Schulden. Da wir keine Schulden haben, enden wir der Einschätzung nach als kreditwürdig (K).

Anhand der Features *Einkommen*, *Vermögen*, *Ist Student* und *Hat Schulden* entscheidet der Decision Tree, ob eine Person kreditwürdig (K) ist oder nicht (N).

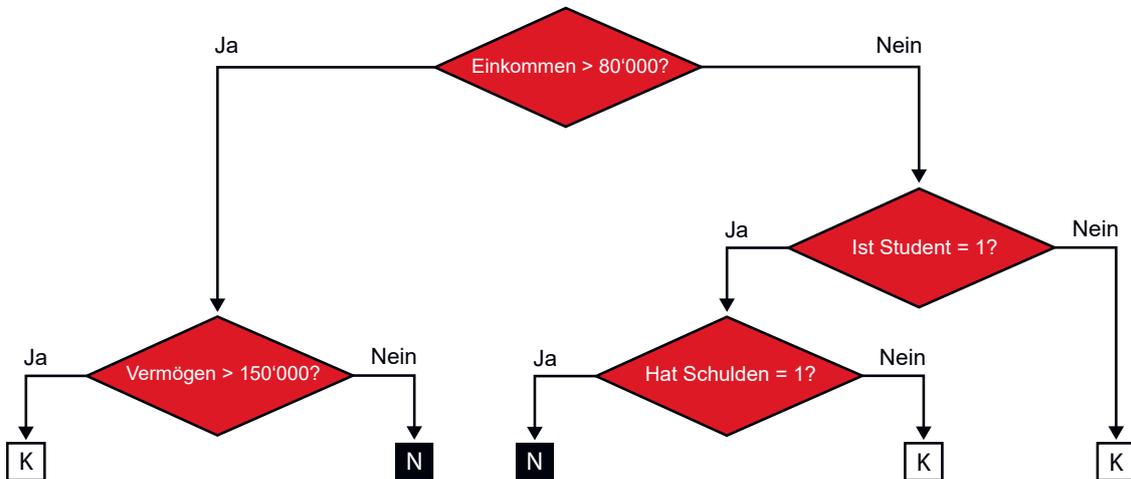


Abbildung 16: Decision Tree

Das Prinzip ist einfach – aber wo kommt hier künstliche Intelligenz ins Spiel? Um dies zu beantworten, stellen wir uns eine andere Frage: Warum wird *Einkommen* als Erstes abgefragt und nicht *Hat Schulden*? Warum der Grenzwert von 80'000 und nicht z. B. 50'000? Warum wird danach entweder nach *Vermögen* oder *Ist Student* gefragt? Genau hier kommt künstliche Intelligenz ins Spiel: Diese Fragen, ihre jeweiligen Grenzwerte und die Ordnung, in der die Fragen aufkommen, wurden im Training anhand der historischen Daten ermittelt, sodass das Modell für den Grossteil der Trainingsdaten die richtige Vorhersage trifft. Mit anderen Worten: Die Fragen und ihre jeweiligen Grenzwerte wurden in einer Weise erarbeitet, welche die Muster in den zugrunde liegenden Daten bestmöglich abbildet.

Das Modell ist also erstmal nur das Gerüst, das durch Training, also durch das Erkennen von Mustern, in eine Form gebracht wird, in der es seiner Funktion, nämlich der Vorhersage oder der Einstufung neuer Daten, nachkommen kann.

Ein Modell ist aber nicht gleich ein Modell – noch vor dem Training müssen wir es konfigurieren. Was gibt es hier aber zu konfigurieren? Um dies zu beantworten, stellen wir uns wiederum eine Frage: Wenn beim Training das Modell wirklich mit Blick auf die Trainingsdaten optimiert wird, wieso gehen wir dann maximal drei Stufen (engl. layers) nach unten? Könnten wir den Decision Tree nicht auf fünf Layer erweitern? Oder 50? Oder vielleicht so viele Layers, wie es Datensätze gibt, sodass wir für jeden einzelnen Datensatz des Trainingsdatensatzes die richtige Antwort bekämen? Das Modell wäre doch dann sicherlich viel besser, oder? Die Antwort finden wir, wenn wir über die Trainingsdaten hinaussähen. Das grenzenlose Zuschneiden (engl. fitting) des Modells auf die Trainingsdaten (und eben nicht auf *allgemeine* Daten ausserhalb der Trainingsdaten) führt zu sogenanntem Overfitting – für die Trainingsdaten gibt das Modell eine perfekte Einschätzung wieder. Für *neue* Daten jedoch fällt es dem Modell dann oft schwer, die richtige Einschätzung zu treffen, weil es nicht die Natur des Problems erfasst hat, sondern stur auf die Perfektion mit Bezug auf die Trainingsdaten hintrainiert hat. Um dies zu verhindern und um das Modell anzuregen, nur die wichtigsten Informationen aus den Trainingsdaten zu nehmen, limitieren wir die maximale Anzahl an Layers.

Diese *maximale Anzahl an Layers* ist ein Parameter, den wir setzen oder *konfigurieren* können. Setzen wir ihn zu niedrig, schafft es der Algorithmus eventuell nicht, die gesamte Logik des Problems abzubilden. Ist er zu hoch, droht *Overfitting*. Wir können noch viele andere Parameter definieren, für die jeweils der optimale Wert

zu finden ist. Das Finden der optimalen Konfiguration, das Optimieren der Parameter, wird auch als *Hyperparameter Optimization* bezeichnet. Dabei analysieren wir verschiedene Kombinationen von Parametern und versuchen, die optimalen Parameter für das beste Modell zu finden.

Wichtige Begriffe: Decision Tree, Random Forest, Xgboost, Logistic Regression, GLM, Neural Networks, Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression, Bayesian Linear Regression, Elastic Net Regression, KMeans, Nearest Neighbor, Overfitting, Hyperparameter Optimization.

Testing

Ein Teil der letzten Aussage des vorhergehenden Abschnitts ist jedoch noch nicht definiert: Was genau bedeutet *beste*, was bedeutet *gut*? Wann ist ein Modell *besser* als das andere? Um diese Fragen zu beantworten, müssen wir zuerst zwei neue Begriffe einführen:

Testdaten: Um das Modell zu evaluieren, müssen wir das Modell auf Daten anwenden (bei denen wir das reale Ergebnis kennen) und das vom Modell berechnete Ergebnis mit dem realen Ergebnis vergleichen. Da wir das Modell jedoch bereits auf den Trainingsdaten trainiert haben, ist es gefährlich, das Modell auch auf den Trainingsdaten zu evaluieren (dies kann man damit vergleichen, Fahrschüler auf der gleichen Strecke mit dem gleichen Verkehr wie in den Übungsstunden zu testen – man prüft nicht, wie sie mit neuen Situationen umgehen können). Stattdessen bietet es sich an, neue Daten zur Hand zu nehmen, um das Modell zu bewerten. Diese Daten bezeichnen wir als Testdaten.

Metrik: Die Metrik ist das Mass, nach dem wir ein Modell bewerten. Wir beschränken uns hier der Einfachheit halber auf Metriken für die Bewertung von binären Klassifikatoren (also Klassifikatoren mit zwei Klassen als mögliches Ergebnis – ohne Beschränkung der Allgemeinheit wird eine Klasse als die zu vorhersehende definiert). Es sei an dieser Stelle erwähnt, dass es natürlich auch für Regression und Clusterung mehrere Metriken gibt, die je nach Problem ein passendes Vergleichsmass darstellen können.

Eine mögliche Metrik kann wie folgt definiert werden: *Unter allen Einschätzungen, welcher Anteil der Einschätzungen war richtig?* Wir können diese Metrik noch anders ausdrücken: Bedienen wir uns erneut dem Beispiel

		Realität	
		Positive	Negative
Einschätzung durch KI	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Abbildung 17: Confusion Matrix

der Kreditwürdigkeit. Die Variable, die wir vorsehen wollen, ist also Ist *kreditwürdig*. Wir bezeichnen Fälle,

- die wir als positiv eingestuft haben und die auch in Wirklichkeit kreditwürdig sind, als **true positives**.
- die wir als negativ eingestuft haben und die auch in Wirklichkeit nicht kreditwürdig sind, als **true negatives**.
- die wir als positiv eingestuft haben, die aber nicht kreditwürdig waren, als **false positives**.
- die wir als negativ eingestuft haben, die aber kreditwürdig waren, als **false negatives**.

Optimalerweise erkennt ein Modell alle positives richtig (true positives) und auch alle negatives (true negatives). Die Zahl der false positives (als positive vorhergesehen, obwohl negative) und false negatives (als negative vorhergesehen, obwohl positive) sollte stets so niedrig wie möglich gehalten werden.

Dann kann die zuvor genannte Metrik auch als $(\text{true positives} + \text{true negatives}) / (\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives})$ definiert werden. Diese Metrik bezeichnen wir auch als Akkurazität (engl. Accuracy). Eine hohe Akkurazität ist allgemein gut – 95 % **Akkurazität** bedeutet schliesslich, dass 95 % aller Daten richtig erkannt worden sind.

95 % muss jedoch nicht in jedem Fall ein Garant für ein *gutes* Modell sein. Nehmen wir beispielsweise an, dass eine Bank nur für 5 % aller Kunden keinen Kredit vergibt. Sie vergibt also im Umkehrschluss an 95 % einen Kredit. Würden wir also das einfachste aller Modelle bauen, nämlich ein Modell, das für jeden Kunden die Einschätzung Ist kreditwürdig wählt, so kämen wir auch hier auf eine Akkurazität von 95 % – das Modell wäre dabei allerdings alles andere als gut: Läge das Ziel der KI nämlich darin, Kunden zu identifizieren, die nicht kreditwürdig sind, so wäre das Ergebnis gänzlich katastrophal.

Eine Alternative, die diese sogenannte Imbalance – also das Ungleichgewicht der beiden Klassen – nicht beeinflusst, umfasst das Paar aus *Präzision* (engl. Precision) und *Recall*. Worum handelt es sich bei diesen beiden Metriken?

- **Präzision:** Die Präzision gibt an, welcher Anteil aller als positiv markierten Fälle tatsächlich positiv ist (also $[\text{true positives}] / [\text{true positives} + \text{false positives}]$).
- **Recall:** Der Recall misst, welcher Anteil der anzunehmenden Fälle wirklich angenommen worden sind (also $[\text{true positives}] / [\text{true positives} + \text{false negatives}]$).

Für beide Metriken ist natürlich ein höherer Wert wünschenswert. Das sogenannte harmonische Mittel aus Präzision und Recall ($2 * \text{Präzision} * \text{Recall} / [\text{Präzision} + \text{Recall}]$) wird auch als **F1 Score** bezeichnet und ist einer der beliebtesten Metriken für Probleme mit unausbalancierten Klassen. Selbst hierbei müssen wir jedoch vorsichtig sein. Wir illustrieren dies an einem Beispiel.

Beispiel: Eine Bank vergibt an 9'500 kreditwürdigen Kunden von 10'000 Interessenten einen Kredit. Betrachten wir die Problematik aus der Sicht, in der uns insbesondere die korrekte Vorhersage von **kreditunwürdigen** Kunden interessiert, wir also **kreditwürdig** als das **positive** Label und nicht kreditwürdige als das negative Label betrachten. Nehmen wir an, das Modell hat **9'600 Kunden** als kreditwürdig kategorisiert (**positives**). Davon waren **9'400** tatsächlich kreditwürdig (**true positive**) und **200** wurden fälschlicherweise als solche kategorisiert (**false positive**). Unter den **400** vom Modell als kreditunwürdig kategorisierten Personas waren **100** kreditwürdige Personen (**false negative**) und **300** kreditunwürdige Personen (**true negative**).

- Die **Akkurazität** wäre hier recht hoch mit $(9'400 + 300) / 10'000 = 97.0 \%$.
- Auch die **Präzision** ist recht hoch mit $9'400 / (9'400 + 100) = 98.9 \%$.

- Der **Recall** liegt etwas tiefer bei $9'400/(9'400+200) = 97.9\%$.
- Der **F1 Score** berechnet sich als harmonisches Mittel von Präzision und Recall zu ca. **98.4%**. Dies ist ein recht hoher Wert. Dies fällt insbesondere dann auf, wenn wir ihn mit einem Modell vergleichen, das für jede Person das Label *kreditwürdig* vergibt – hier liegt der F1 Score bei **97.4%**. Auch wenn wir für jede Person zufällig mit 95%iger Wahrscheinlichkeit auf kreditwürdig entscheiden würden, stünden wir mit einem F1 Score von **95.0%** nicht besser da.

Wie sieht das Ganze aber aus, wenn unser Ziel darin besteht, möglichst viele der kreditunwürdigen Kunden aus dem Verkehr zu ziehen? In diesem Fall definieren wir **kreditunwürdig** als das **positive Label** (also als das Label von Interesse).

- Die **Akkurazität** liegt weiterhin bei $(9'400 + 300)/10'000 = 97.0\%$ – also immer noch alles gut.
- Die **Präzision** liegt bei $300/(300 + 200) = 60.0\%$ – schon deutlich tiefer als die zuvor berechnete Präzision.
- Auch der **Recall** liegt nur noch bei $300/(300 + 100) = 75.0\%$.
- Der **F1 Score** liegt auch nur noch bei **66.7%**.

Dieses Ergebnis ist wiederum besser als stures Raten auf Kreditunwürdigkeit oder Kreditwürdigkeit. Ob das Ergebnis jedoch gut genug ist, bleibt zu evaluieren und liegt am Ende des Tages an den Zielen der Bank.

Das obige Beispiel hat gezeigt, dass der F1 Score bzw. die F1 Scores eine gute Metrik für die Bewertung eines Modells darstellt. Metriken stellen jedoch nicht nur eine Möglichkeit dar, ein Modell zu bewerten. Vielmehr erlauben sie es, mehrere Modelle eben anhand dieser Metrik zu verwerfen. Vor allem wenn es um die Konfiguration eines Modells geht, hilft die Metrik dabei, die optimale Konfiguration zu finden, denn auch hier gilt: Unter zwei Modellen desselben Modelltyps mit verschiedenen Konfigurationen ist das Modell mit der besseren Metrik im Regelfall vorzuziehen – seine Konfiguration ist besser als die des anderen Modells.

Betrachten wir nochmals das Beispiel aus dem vorigen Kapitel, in welchem wir einen Decision Tree betrachtet haben. Woher wissen wir, ob wir für den Decision Tree 5, 50, 500 oder 5000 Layers als maximale Tiefe auswählen sollen? A priori ist dies schwer zu bewerten – wir müssen für jede einzelne Konfiguration (5, 50, 500 oder 5000 Layers) das Modell trainieren und die verschiedenen Modelle anhand einer Metrik, etwa dem F1 Score, vergleichen. Diese Art der Optimierung der Konfiguration, die wir bereits im letzten Kapitel vorgestellt haben, nennt sich Hyperparameter Optimization.

Nachdem wir den technischen Abschnitt der Erstellung einer KI beleuchtet haben, würden wir noch gerne auf den *Cross Industry Standard Process for Data Mining* (CRISP-DM) verweisen. Dieser häufig (bewusst oder unbewusst) verwendete Standard umfasst die folgenden Schritte:

1. **Business Understanding:** Fachliches Verständnis eines Problems
2. **Data Understanding:** Grundlegendes Verständnis der Daten
3. **Data Preparation:** Das Modifizieren und Anreichern der Daten
4. **Modeling:** Das Erstellen des Modells, inkl. Training
5. **Evaluation:** Die Modell-Evaluierung anhand fachlicher und technischer KPIs
6. **Deployment:** Das Produktivsetzen des Modells, z. B. lokal oder auf der Cloud

Sich diesen Standard regelmässig wieder zurück ins Gedächtnis zu rufen, ist aus zwei Gründen vorteilhaft: Zum einen betont dieser Ansatz, dass das Erstellen einer KI ohne fachliches Verständnis wenig sinnvoll ist. Auch wenn sich ein Data Scientist in seiner Domain wohlfühlt: In jedem Projekt, das auf einen Mehrwert durch KI zählt, sollte er sich zumindest ein grundlegendes Verständnis des fachlichen Hintergrunds aneignen. Feature Engineering ist ein wichtiger Teil im Pre-Processing und je besser das Fachverständnis, umso leichter das Generieren neuer Features, welche die Mustererkennung für das Modell leichter machen. Zum anderen betont dieser

Vorgang auch, dass ein Projekt mit Ziel einer KI nicht mit Hyperparameter Optimization bzw. dem Trainieren des optimalen Modells aufhört. Die Inbetriebnahme resp. das Produktivsetzen des Modells, die Einbettung des Modells in bestehende Prozesse, die Ablösung von Prozessen, etwaige grafische Oberflächen, die das Ergebnis visualisieren, sind genau zu durchdenken und nicht zu unterschätzen.

Wichtige Begriffe: Testing Daten, Metrik, Akkurazität, Präzision, Recall, F1 Score, True Positives, True Negatives, False Positives, False Negatives.

11.5 Anfälligkeiten eines KI-Modells

Bei der Realisierung einer KI muss man sich zu jeder Zeit der Limitationen eines KI-Modells bewusst sein. Ein solches Modell ist höchstens so gut wie die Daten, die es bekommt. Wenn die Daten, mit denen eine KI trainiert wird, nicht die richtige Form besitzen oder anderweitig eine niedrige Qualität aufweisen, können die Parameter noch so lange konfiguriert werden – ein gutes Ergebnis wird nicht mehr erzielt. Doch was sind die Anfälligkeiten einer KI?

Overfitting: Wir haben bereits den Begriff des Overfittings eingeführt, bei dem sich ein Modell durch Training zu stark an den Trainingsdaten orientiert. In diesem Fall ist die Performance auf den Trainingsdaten generell sehr hoch, gleichzeitig jedoch sehr niedrig auf den Testdaten. Overfitting kann umgangen werden, indem die Parameter so gewählt werden, dass das Modell um eine hinreichende Verallgemeinerung des Problems nicht umhinkommt. Hyperparameter Optimization hilft uns dabei, ebendiese Parameterkonfiguration zu finden. Indem die Performance des Modells nicht auf den Trainingsdaten, sondern vielmehr auf den Testdaten evaluiert wird, schafft es Hyperparameter Optimization, die passenden Parameter zu finden, die eine hinreichende Verallgemeinerung über alle Daten hinweg erlauben.

Allerdings birgt selbst hierin Gefahr. Damit sich das Modell nicht auf die konstant bleibenden Testdaten einspielt, verwendet man in diesem Prozess zumeist die sogenannte Cross-Validation. Die Idee von Cross Validation ist es, die zur Verfügung stehenden Daten in eine Anzahl – wir nennen sie hier mal n – gleichgrosser Gruppen zu unterteilen. Nun wird für jede Modellevaluierung n -mal trainiert, und zwar immer auf $n-1$ -Gruppen, sodass die jeweils verbleibende Gruppe als Testdaten fungiert. Die Güte eines Modells, bemessen zum Beispiel in Form des F1 Score, wird dann nicht mehr nur auf einem einzelnen Training bewertet, sondern auf den n -Ergebnissen der n -Trainingsdurchgänge. Ein wichtiger Schritt weg vom Overfitting.

Bias: Bias oder Voreingenommenheit bezeichnet den Zustand, wenn KI bestimmte Attribute «unfair» behandelt. Unfair ist dabei so zu verstehen, dass Datensätze, die ein bestimmtes Attribut erfüllen, vermehrt durch das Modell mit einer Klasse versehen werden, obwohl diese Einstellung in vielen Fällen falsch ist.

Nehmen wir als Beispiel die Kategorisierung der Kreditwürdigkeit. In der Bewerbung geben wir an, in einer Postleitzahl zu leben, in der sonst eine hohe Dichte an Erstsemester-Studenten lebt, deren Bewerbung abgelehnt worden ist. Das Modell wurde unter anderem auf diesen Daten trainiert. Obwohl wir einen stabilen Job haben und schon ein kleines Vermögen angespart haben, wird uns die Kreditwürdigkeit verweigert. Einzig wegen der Postleitzahl und der Fehleinschätzung des Systems, beruhend auf eben diesem einen Attribut.

Bias kommt in vielen verschiedenen Formen vor. Oftmals ist Bias das Resultat unausgewogener Trainingsdaten. Da ein jedes System das Ziel der maximalen Fairness als Ziel haben sollte, ist die Zusammensetzung der Trainingsdaten sorgfältig durchzuführen. Nicht nur die Zusammensetzung der Trainingsdaten kann dabei helfen. Gewisse Features sollten, falls nicht unbedingt notwendig, nicht miteinbezogen werden. Dazu gehören etwa Alter, Geschlecht und Herkunft.

Intransparenz: Künstliche Intelligenz schafft es, komplexe Sachverhalte anhand von historischen Daten zu

lernen und anzuwenden. Die Vorhersage alleine reicht jedoch in vielen Fällen nicht. Entscheidungen können nicht immer nur aufgrund einer einzigen Einschätzung getroffen werden, sondern müssen begründet werden. Eine Einsicht darin, wieso sich Algorithmen für (oder gegen) eine spezifische Vorhersage entscheiden, liefert das sogenannte Explainable AI (engl. für beschreibende KI).

Auf der linken Seite sind die Features zu sehen, welche – für diesen Fall – gegen eine Kreditwürdigkeit sprechen,

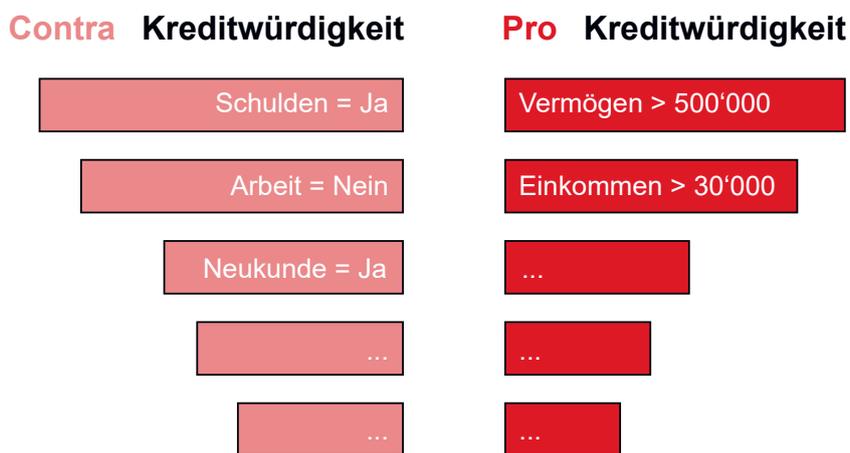


Abbildung 18: Illustration eines Explainable AI für einen Fall

während rechts die Features abgebildet sind, die für eine Kreditwürdigkeit sprechen. Die Länge der Balken hängt dabei mit der Relevanz des Pro/Contra-Features zusammen. In diesem Beispiel scheint – zumindest für die Top-5-Pro/Contra – Contra zu überwiegen.

Concept Drift: Die Muster, die beispielsweise im Bereich des überwachten Lernens von Inputdaten auf das Label schließen lassen, sind nicht immer konstant. Wir bedienen uns erneut des Beispiels der Kreditwürdigkeit. Nehmen wir an, unser Modell ist auf Daten trainiert, die zu einer Zeit entstanden, als Bänker mit sehr konservativen Regeln auf Kreditwürdigkeit geschlossen haben. Das Modell lernt dann genau diesen konservativen Ansatz. Beschliesst also die Bank, einen neuen Kurs einzuschlagen und etwa das Alter nicht mehr in Betracht zu ziehen, so wird diese Kursänderung so lange nicht durch das Modell abgebildet sein, bis die neuen Daten dem Modell geliefert werden (und optimalerweise in dieser Hinsicht veraltete Daten dem Modell vorenthalten werden).

Wichtige Begriffe: Overfitting, Bias, Explainable AI, Concept Drift.

11.6 Der schnelle Weg zum ersten Modell

Dieser kurze Exkurs über künstliche Intelligenz gibt nur einen kleinen Einblick in die verschiedenen Bereiche und die Möglichkeiten, die uns geboten sind. Die Entwicklung der letzten Jahre hat zwei Bilder geprüft. Zum einen hielt und hält KI mehr und mehr Einfluss in unserem Alltag. Zum anderen ist jedoch auch der Hype um künstliche Intelligenz relativiert worden.

Wenn wir also in unserem alltäglichen Arbeitsleben Prozesse erkennen, die gewisse Muster aufzeigen, welche sich eine künstliche Intelligenz aneignen könnten, sollte man sich auf jeden Fall Gedanken machen, den Einsatz von KI zu prüfen. Diese Prüfung kann relativ schnell stattfinden – mithilfe sogenannter AutoML-Ansätze, welche sowohl on premise (h2o, TPOT, autosklearn) als auch auf der Cloud (Google Cloud AutoML, Azure Auto ML, Amazon SageMaker Autopilot) angeboten werden, muss der User kaum mehr als das Problem definieren.

Hyperparameter Optimization und andere Anpassungen werden durch die AutoML-Algorithmen übernommen.

Gleichzeitig ist es aber wichtig, sich genau zu überlegen, welchen Mehrwert eine solche KI mit sich bringen kann. Ein weiterer zentraler Punkt ist zudem, wie sich die neue KI in die Prozesse einbetten kann, ob die Prozesse mit Hintergrund der neuen KI angepasst oder sogar ganz abgelöst bzw. erneuert werden können.

Befinden wir uns also in einem solchen Projekt, bei dem wir Vorteile durch die automatische Erkennung von Mustern und somit den Einsatz von KI ausgemacht haben, so sollten wir uns zu jeder Zeit wieder Standards wie den CRISP-DM oder ähnliche Prozesse vor Augen halten. Ohne Sie, die zukünftigen Data Scientists, wird sich das Projekt schwertun, diese Muster zu erlernen und in der Produktion zu erkennen. Ohne Fachexperten und Fachverantwortliche werden Sie sich in der Regel jedoch schwertun, Ihr Modell zu evaluieren, zu optimieren, auf das tatsächliche Problem auszurichten und passend in bestehende Prozesse einzubetten bzw. diese abzulösen und neu zu definieren.

Die Interdisziplinarität ist ein zentraler Bestandteil eines erfolgreichen KI-Projekts, die sie von Beginn bis Abschluss des Projekts pflegen und leben sollten. Schlussendlich ist ein KI-Projekt nicht eindimensional. Es lebt vom stetigen Feedback des Fachs und die Übersetzung der fachlichen Anforderungen in neue technische Features durch Sie, die Data Scientists.

Wichtige Begriffe: AutoML, CRISP-DM.

12 Digitale Transformation und neue Rollenbilder

Die Digitalisierung ist allgegenwärtig und als «Megatrend» bereits tief im Business verankert. Sich dieser zu entziehen, ist also weder im privaten noch im geschäftlichen Umfeld bereits seit Jahren kaum noch möglich und in vielerlei Hinsicht über alle Generationen hinweg mehrheitlich nicht (mehr) gewollt; zu gross ist deren Durchdringung und der sich daraus ergebende Nutzen. Natürlich gibt es neben den zum Teil unklaren Verhältnissen rund um den Datenschutz und die Datensicherheit in anderen Bereichen der Digitalisierung fragwürdige oder teils gar absurde Ausprägungen. Aber davon mal abgesehen, zeigt die sich seit Jahren oder gar Jahrzehnten stetig ausbreitende und mittlerweile grossflächig vorhandene Akzeptanz der Digitalisierung im Allgemeinen, dass diese über alle Generationen hinweg durchaus seine Berechtigung hat. Doch was bedeutet eigentlich «Digitalisierung»? Wie spielen «Digitalisierung» und «Digitale Transformation» zusammen? Und welche neuen Rollenbilder werden in der digitalen Finanzwelt in Zukunft an Bedeutung gewinnen? Diese und weitere Fragen wird das nachfolgende Kapitel in der nötigen Tiefe theoretisch durchleuchten und punktuell mit praxistauglichen Beispielen und Grafiken erläutern.

12.1 Digitalisierung

Vom Ursprung her bedeutet Digitalisierung nichts anderes als das Umwandeln analoger Werte in digitale Formate und deren anschliessende Verarbeitung und Speicherung auf einem digitalen Medium. Unter Digitalisierung sind also Technologien zu verstehen, die uns als Privatpersonen und Unternehmen in unterschiedlichen Bereichen und in unterschiedlicher Ausprägung durchdringen und dadurch das Verhalten, die Prozesse, die Kollaboration, die Wettbewerbsfähigkeit und vieles andere mehr grundlegend beeinflussen.

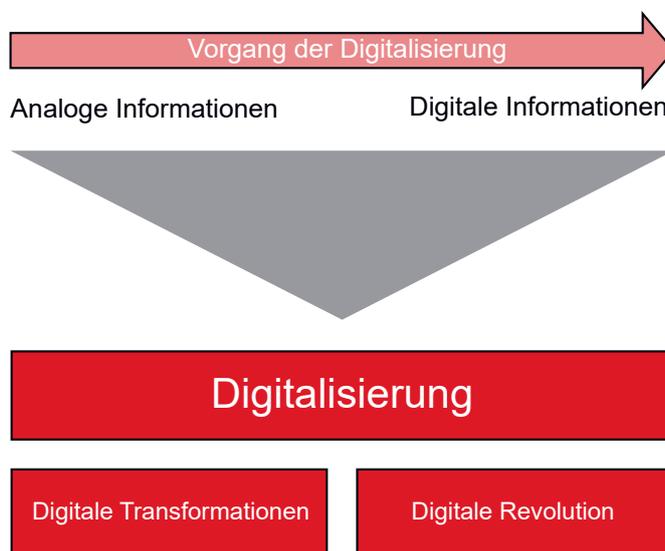


Abbildung 19: Digitale Transformation als Teil der Digitalisierung

Die Digitalisierung schafft also durch technologische Umstellung bestehender und die Schaffung neuer Produkte die notwendigen Voraussetzungen für die sich daraus als Folge ergebende digitale Transformation.

12.2 Zielsetzungen einer digitalen Transformation

Wie es der Name bereits erahnen lässt, geht es bei der digitalen Transformation um einen durch die Digitalisierung ausgelösten Prozess des Wandels. Treiber dieses Wandels können dabei vielfältiger Natur sein, wobei mit zunehmender Menge an unterschiedlichen Treibern bei gleichzeitiger Erhöhung des Realisierungsdrucks die Geschwindigkeit der Transformation immer mehr zunimmt.

Wesentliche Treiber einer digitalen Transformation können unter anderem folgende sein:

Externe Treiber	Interne Treiber
Wirtschaft	Fähigkeiten der Mitarbeitenden und deren Arbeitsverhalten
Konkurrenz	Unternehmensstrategie
Kunden	Forschung und Entwicklung
Ökologie	etc.
etc.	

Um ein besseres Verständnis für die digitale Transformation zu entwickeln, kann als einfache Hilfestellung das **Modell der drei Stufen des digitalen Wissens** (Martin, 2008) beigezogen werden. Das Modell unterteilt den Weg hin zum eigentlichen Transformationsprozess in drei Schlüsselstufen und beschreibt diese so, dass zuerst die digitale Kompetenz aufgebaut werden muss (Level 1: Digital Competence – skills, concepts, approaches, attitudes etc.), um mit den neuen Technologien umgehen zu können. Anschliessend folgt auf der zweiten Stufe (Level 2: Digital Usage – professional/discipline application) die innovative und kreative Nutzung sinnvoller und durchdachter digitaler Anwendungen, die in Anlehnung an die übergeordnete Strategie als Arbeitsmittel zur Verfügung gestellt werden. Erst auf der dritten Stufe folgt die eigentliche digitale Transformation.

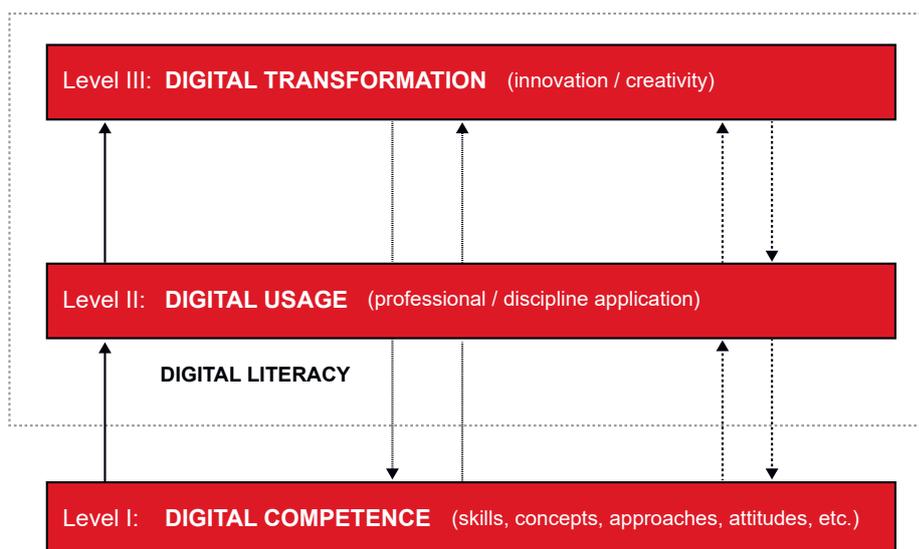


Abbildung 20: Levels of Digital Literacy (Martin, 2008)

Mit Erreichen des dritten Levels werden im Rahmen der eigentlichen digitalen Transformation (Ruoss, 2015) die folgenden Zielsetzungen in den Kernbereichen «Erhöhung der Kundenbindung», «Interne Effizienzsteigerung» und «Umsatzsteigerung durch neue Produkte» verfolgt.

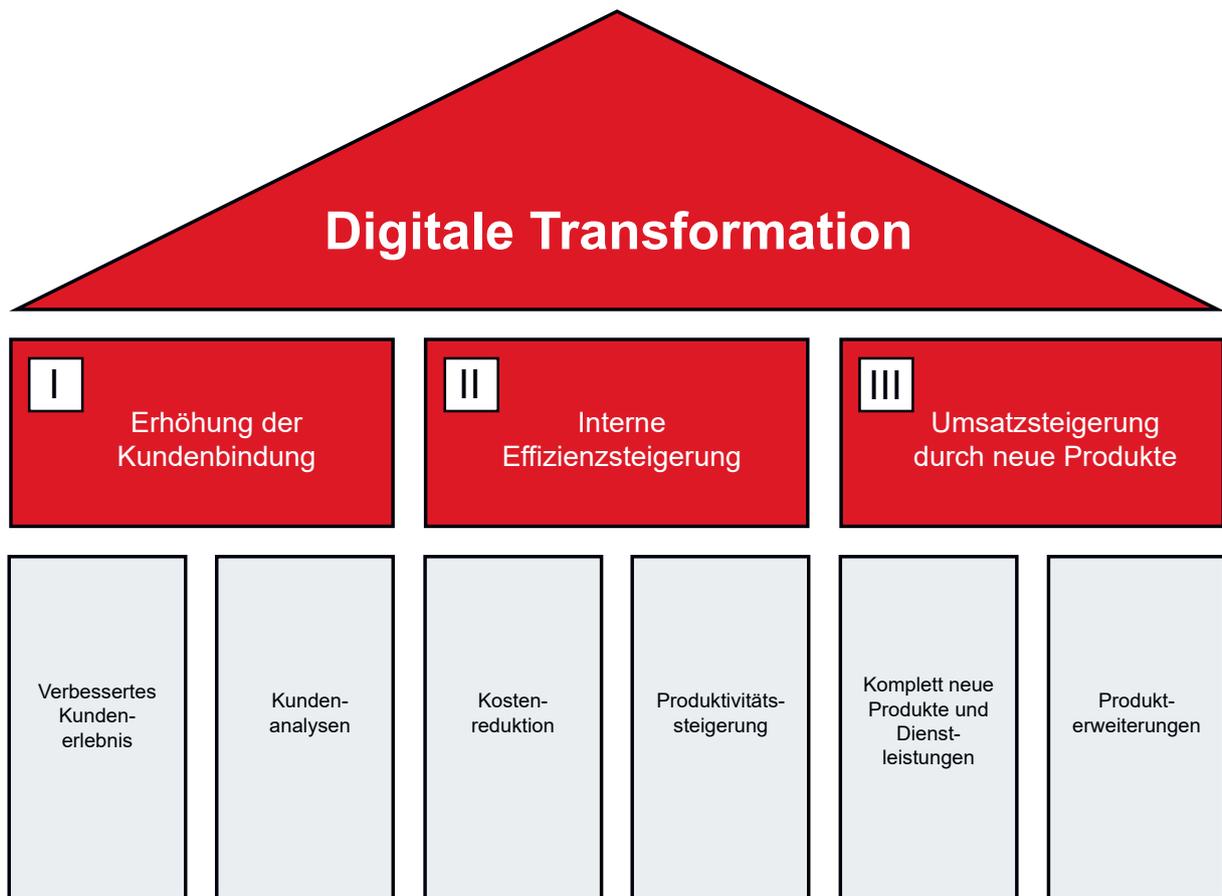


Abbildung 21: Ziele der digitalen Transformation (Ruoss, 2015)

Erhöhung der Kundenbindung

Das kollaborative Verhalten der Menschen und wie diese die für sie relevanten Informationen im Privatbereich und/oder im Geschäftsleben zusammentragen, hat sich bis heute aufgrund der durch die Digitalisierung geschaffenen Möglichkeiten bereits stark verändert und wird das in Zukunft auch weiter tun; mit einer erwartungsgemäss stetig zunehmenden Geschwindigkeit. So erfolgt beispielsweise die Kundenansprache – nicht zuletzt aufgrund der unterschiedlich ausgeprägten Kundenerwartungen – über verschiedene (Social-Media-) Kanäle und digitale Plattformen, wodurch sich die Komplexität für Unternehmen stetig erhöht. Sämtliche Prozesse rund um das Kundenerlebnis sind unternehmenseitig auf die Kundenbedürfnisse auszurichten. Genauso komplex gestaltet sich auf der anderen Seite die Informationsbeschaffung durch die Unternehmen, um aktuelle und zukünftige Kundenbedürfnisse in Erfahrung zu bringen und Produkte und Dienstleistungen entsprechend laufend darauf auszurichten.

Unternehmensinterne Effizienzsteigerung

Die unternehmensinterne Effizienzsteigerung ergibt sich aus einer durch die Digitalisierung erreichten Kostenreduktion bei gleichzeitiger Steigerung der Produktivität. Die Digitalisierung bedingt zwar eine initiale Investition und löst anschliessend möglicherweise laufende Kosten aus, kann aber Unternehmen durch die Automatisierung von Prozessen einen im Verhältnis zu den Investitionsausgaben ungleich höheren Nutzen in Form von Kostenreduktion stiften. Gleiches gilt für die Produktivitätssteigerung. Wenn also ein Unternehmen einerseits durch die Digitalisierung die Kosten reduzieren kann und andererseits damit produktiver wird, resultiert daraus eine maximale Rendite in Form von gesteigerter Effizienz, was sich letztendlich wiederum auf die Liquidität und das Ergebnis eines Unternehmens auswirkt.

Umsatzsteigerung durch neue Produkte

Digitalisierung ist ein wesentlicher Innovationstreiber. Je nach Art und Ausprägung der Innovation können mit bzw. aus der Digitalisierung heraus komplett neue Produkte und Dienstleistungen entstehen oder bestehende, je nach bereits etablierten Produkten, durch die Digitalisierung erweitert und damit neue Märkte erschlossen werden.

Wichtig ist an dieser Stelle zu erwähnen, dass die digitale Transformation nur dann erfolgreich umgesetzt werden kann, wenn alle der vorerwähnten Kernbereiche eines Unternehmens sinnvoll zusammenspielen. Die digitale Transformation scheitert folglich dann, wenn die Voraussetzungen aus den Level I und II nicht gegeben sind und/oder die Kernbereiche der digitalen Transformation nicht aufeinander abgestimmt sind. Ausserdem führt die digitale Transformation automatisch dazu, dass sich ein Unternehmen intensiv mit dem eigenen Geschäftsmodell und den entsprechenden operativen Prozessen auseinandersetzen muss.

12.3 Herausforderungen der digitalen Transformation

Die Herausforderungen entstehen bei der digitalen Transformation (Ruoss, 2015) aus den drei Bereichen «Initiierung», «Implementierung» und «Koordination»:

Initiierung

Fehlende Impulse, regulatorische Zwänge, Sicherheits- und Datenschutzbedenken, Reputation, unklarer Business Case

Implementierung

Fehlendes Know-how, fehlende Veränderungskultur, IT-Schwierigkeiten

Koordination

Silodenken, mangelhafte übergreifende Koordination, unzureichende «Management Attention» und Top-down-Führung

12.4 Roadmap einer Digital Finance Transformation

Wichtige Voraussetzung für eine strukturierte Digital Finance Transformation ist nebst einer begünstigenden Unternehmenskultur eine überlegte und klar definierte Strategie, die auch eine klare Haltung zur digitalen Vision des Unternehmens einnimmt. Sind diese übergeordneten Voraussetzungen gegeben, gelingt die operative Umsetzung am besten mit einer Roadmap (ExpertFocus, 2019/12), welche die wesentlichen Meilensteine vorgibt. Der gesamte Prozess wird über alle Meilensteine hinweg durch eine gegenüber den Mitarbeitenden laufende, transparente Kommunikation begleitet. Damit wird der digitale Wandel den Mitarbeitenden leichter zugänglich gemacht und allfällige Unsicherheiten oder Ängste werden abgebaut.

Die Rangfolge der Meilensteine folgt einer gewissen Logik. Die erfolgreiche Umsetzung der digitalen Transformation ist dann gewährleistet, wenn jeder Schritt der Reihe nach vollständig und erfolgreich umgesetzt ist.

Elektronische Archivierung

Bevor mit der eigentlichen Automatisierung gestartet werden kann, muss zuerst das physische Archiv in ein digitales überführt werden. Dabei sind die unternehmensinternen und regulatorischen Vorgaben zu beachten. Ein weiterer Effizienzgewinn wird dadurch erreicht, indem mit den Erzeugern von Dokumenten wie Rechnungen und sonstigen Belegen vereinbart wird, diese nur noch in elektronischer Form auszutauschen.

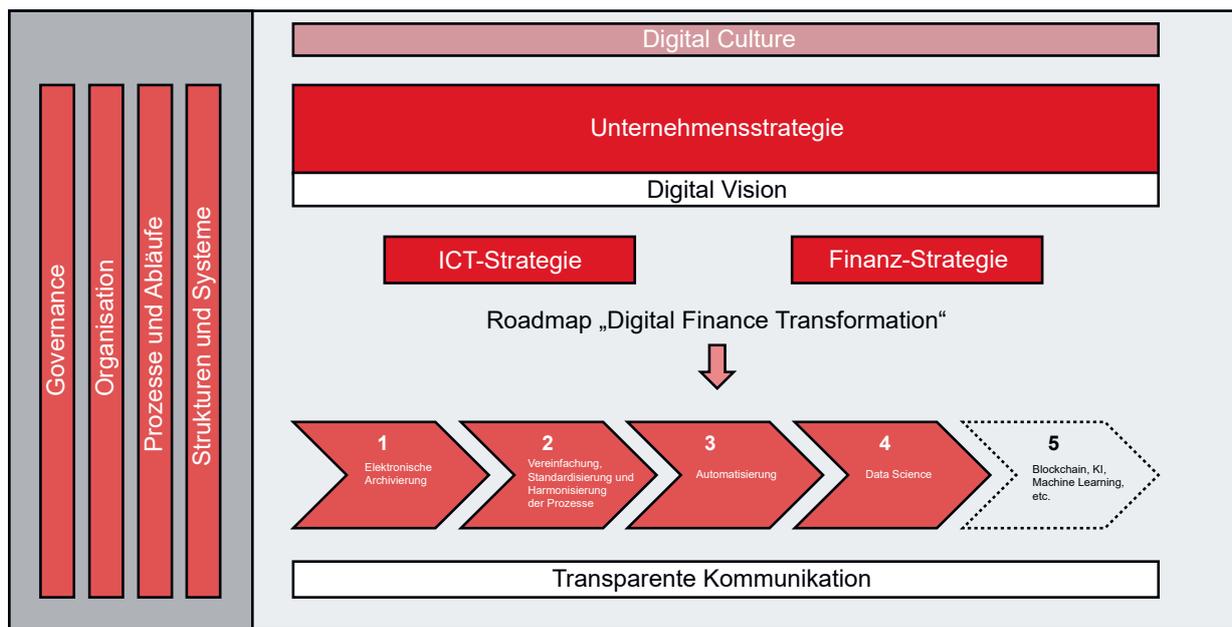


Abbildung 22: Digital Finance Transformation

In diesem Fall sollte sich ein Unternehmen für die Umsetzung der elektronischen Archivierung insbesondere die nachfolgenden Fragen stellen (ExpertFocus, 2019/12):

- Wie sieht die momentane Archivierungs-Politik aus? Wie muss diese an das Digitalisierungszeitalter angepasst werden und welche unternehmensinternen Stellen müssen dabei involviert werden?
- Welches zentrale System wird für die elektronische Archivierung benutzt?
- Wo werden die Daten gespeichert?
- Wie ist es möglich, sämtliche eingehenden Dokumente in eine digitale Form zu bringen?

Vereinfachung, Standardisierung und Harmonisierung der Prozesse

Im zweiten Schritt werden innerhalb eines zentralen Enterprise-Resource-Planning-Systems (ERP) einfache Prozesse zur Effizienzsteigerung und Fehlerreduktion standardisiert und in programmierte Routinen transferiert. Durch die Vereinfachung, Standardisierung und Harmonisierung können im nächsten Schritt insbesondere einfache End-to-End-Prozesse automatisiert werden.

Automatisierung

Die Automatisierung der Prozesse kann auf unterschiedliche Arten erfolgen. Ganz generell geht es dabei darum, dass bisher manuell abgearbeitete End-to-End-Prozesse neu ohne menschliche Einflussnahme automatisch verarbeitet werden. Hinsichtlich der Technologie kann die Automatisierung durch die Möglichkeiten des zentralen ERP-Systems und/oder durch den Einsatz von Robotic Automation (RPA) oder künstlicher Intelligenz erfolgen, wobei davon auszugehen ist, dass die ERP-Systeme in Zukunft alle Möglichkeiten der Automatisierung vereinen werden.

Data Science

Ist das betriebliche Rechnungswesen vollständig automatisiert, folgt die Königsdisziplin, die Data Science. Dabei geht es darum, aus den nun vollständig digital vorhandenen bisherigen und laufend neu dazukommenden Daten durch Data Science-Technologien Erkenntnisse in Form von Muster zu gewinnen und diese nutzenstiftend im Unternehmen weiter zu verwenden. Daraus gewonnene Erkenntnisse können anschliessend für die Entwicklung neuer oder die Erweiterung bestehender Produkte und Dienstleistungen genutzt werden.

Data Science beschränkt sich aber nicht nur auf Erkenntnisgewinne durch Datenauswertung aus bestehenden internen Systemen. Sie geht viel weiter und kombiniert die internen Daten mit extern verfügbaren Daten aus unterschiedlichen Kanälen. Nur so können die umfassenden Bedürfnisse aller relevanten Stakeholder erhoben und nutzenstiftend auf die eigenen Produkte und Prozesse adaptiert werden.

Blockchain, KI, Machine Learning

Der letzte Schritt der Roadmap bezieht sich auf die ständige und weitere Verbesserung von Prozessen und Datenauswertungen durch Instrumentarien wie das Process Mining sowie den Einsatz von in Zukunft denkbaren neuen Technologien.

Sämtliche Schritte der Roadmap sind Bestandteil eines agil aufgesetzten Controlling-Prozesses und sollten laufend auf technologische Änderungen oder neue Anforderungen überprüft werden. So gesehen ist die digitale Transformation kein einmaliger statischer Prozess, sondern ist selbst nach erfolgreicher Umsetzung einer digitalen Transformation als laufender Controlling-Prozess zu implementieren.

13 Veränderung der finanziellen Führung und Steuerung der Organisation

Durch das mit der Data Science in Verbindung stehende Data Mining (automatisierte Auswertung grosser Datenmengen zur Erkennung bestimmter Muster) sind Unternehmen in der Lage, aus intern und extern gesammelten und miteinander sinnvoll kombinierten Daten Muster zu erkennen, Szenarien zu entwickeln und letztendlich Entscheidungen zu fällen (Data Driven Decision Making). Die Herausforderung des Controllings und letztendlich auch der finanziellen Führung besteht darin, jederzeit die relevanten Daten aus den entsprechend relevanten in- und externen Quellen zusammenzubringen, diese auszuwerten und daraus die strategisch und operativ relevanten Schlüsse zu ziehen. Die finanzielle Führung erstreckt sich also über alle Kanäle hinweg bis hin zu Auswertungen digitalisierter Produkte und erhöht damit Anforderungen an die Funktion des Controllers.

13.1 Neue Rollenbilder in der digitalen Welt der Finanzfunktionen

Einführung und Rollenverständnis

Das heute noch weitverbreitete Verständnis von Führung bezieht sich auf die Positionsmacht von Führungskräften. Führungskräfte waren durch ihre Funktion bemächtigt oder mindestens beauftragt, was dazu führte, dass sie loben, sanktionieren oder bestrafen konnten. Vor einigen Generationen durfte die Autorität von Führungskräften nie angezweifelt werden, weil diese Kraft ihrer «Bemächtigung» immer recht hat. Im neuen Führungsverständnis werden Führungskräfte nicht mehr mit Macht, sondern mit Aufgaben und Rollen ausgestattet. Heute wissen Führungskräfte nicht mehr alles und sind nicht mehr unfehlbar – im Gegenteil, sie machen auch Fehler. Eigene Fehler einzugestehen, auf Fehlentscheidungen zurückzukommen und diese zu korrigieren, wird heute von den allermeisten Mitarbeitenden als Stärke wahrgenommen. Der kritischen Selbstreflexion von Führungskräften kommt eine vorbildhafte Wirkung zu. Etablierte Führungskräfte sprechen heute davon, dass sie in ihrer Funktion den Mitarbeitenden der Teams dienen. Damit ist die Führungskraft, ob hierarchisch legitimiert oder nicht, auch Teamentwickler.

Organisationsexperten sind sich heute einig, dass Sparten- und Matrixorganisationen operativ nicht überlebensfähig sind, weil sie dynamischen Märkten nicht standhalten können. Anstelle der hierarchischen Organisationen treten modulare und dezentrale Netzwerkorganisationen.

Das Rollenverständnis der Mitarbeitenden basiert auf Aufgaben, Verantwortung sowie Rechten und Kompetenzen. Der Mitarbeitende hat das Recht (und die Pflicht), sich innerhalb der Rolle frei zu bewegen, seine Aufgaben zu interpretieren, die Aufgabe auszuführen und wo notwendig zu optimieren.

Die **Position** beschreibt den formalen Platz, den jemand in einer Organisation einnimmt, z.B. Verkaufsleiter, CEO, Projektleiter, Teamleiter. In der **Funktion** sind Zweck und Ziele der Tätigkeit definiert. In der **Rolle** sind Aufgaben, Rechte, Pflichten und Kompetenzen umschrieben. Übergeordnet sind dies auch Auszüge aus Prozessbeschreibungen der Organisation. Alle Rollenbeschreibungen zusammen beschreiben die gesamte Organisation und alle Aufgaben.

Dabei sind die folgenden Punkte zu beachten:

- Rollenbeschreibungen sind nicht personengebunden
- Eine Person kann mehrere Rollen ausfüllen
- Rollen können auch nur zeitweise besetzt sein

Die beschriebenen Rollen sind nicht frei interpretierbar, sondern durch die Pflichten und Kompetenzen begrenzt. Rollenkonzepte setzen voraus, dass niemand über den anderen gestellt ist, sondern alle in der Organisation

gleichwertig sind. In der Rolle sind die Aufgaben beschrieben. Es ist damit auch wahrscheinlicher, dass kompetente und potenzielle Mitarbeitende sich für eine neue oder andere Rolle prädestinieren oder gar aufdrängen.

Das bisherige Führungsverständnis gerät unter Druck. Gefragt sind nicht mehr «Vorgesetzte», sondern «Führungskräfte», die in der Erfüllung ihrer Rolle die Potenziale von Mitarbeitenden abrufen und fördern sowie den Teams Hindernisse aus dem Weg räumen.

Der Vorgesetzte wird durch seine Mitarbeitenden nicht mehr mit Ergebnissen bedient, sondern dient als Führungskraft seinen Mitarbeitenden und Teams, indem er deren Potenziale abrufen.

Diese Überlegungen haben auch eine Auswirkung auf das Personalmanagement. Unternehmen legten in der Vergangenheit viel Wert auf die harmonische Kompatibilität von Bewerbern zu Team und Organisation. Heute geht es zunehmend darum, neben sozialen und kulturellen Überlegungen auch Innovationen zu fördern und Teams zu Höchstleistungen anzutreiben. Dies erfordert Reibung und Perspektivenreichtum. Zielsetzung der Mitarbeiterbeschaffung ist es nicht mehr, mehr vom «Gleichen» zu suchen, sondern mehr vom «Anderen». Damit haben viele etablierte Vorgesetzte ein Problem, schliesslich sind sie diejenigen, die sich bei Meinungsverschiedenheiten in der Rolle der Entscheider sehen. Dies ist insofern falsch, weil Mitarbeitende immer mehr in der Lage sind, Konflikte auszudiskutieren und unternehmerische Lösungen zu suchen.

Rollenbilder im Zeitalter der Digitalisierung

Die sich aus der Digitalisierung ergebenden neuen Rollenbilder erfordern vermehrt analytische und diagnostische Fähigkeiten in Bezug auf mechanische und elektronische Komponenten und Funktionen. In diesem Sinne nimmt die Komplexität unserer Arbeit zu. IT-Tools, Roboter und Prozessoren werden immer tiefer in unser Leben und unsere beruflichen Kompetenzen vordringen und diese verändern. Während diese Systeme früher eher einfache Tätigkeiten bewerkstelligen konnten, sind sie zunehmend in der Lage, komplexere und wiederkehrende Aufgaben zu übernehmen. Die Analyse der neuen Anforderungen an die Arbeitswelt zeigen, dass zwar viele einfache Arbeiten und Tätigkeiten entfallen werden, sich jedoch gleichzeitig die bestehenden Berufsbilder verändern sowie neue Tätigkeiten und Berufsbilder entstehen werden.

VUCA-Welt

Mit den Vorteilen der Globalisierung hat unsere Gesellschaft auch einige Nachteile erfahren. Während in den vergangenen Generationen ausschliesslich Naturkatastrophen und Krankheiten globale Auswirkungen hatten, zeigte beispielsweise der Zusammenbruch des Ostblocks zu Beginn der 1990er-Jahre, dass Veränderungen politischer Systeme globale Auswirkungen auf Volkswirtschaft und Gesellschaft haben.

Als Akronym der multilateralen Welt wurde der Begriff VUCA geschaffen. Dieser steht für die Begriffe Volatility, Uncertainty, Complexity und Ambiguity (dt. Volatilität = Unbeständigkeit, Unsicherheit, Komplexität und Mehrdeutigkeit).

Auf Unternehmensebene zeichnet sich die VUCA-Welt dadurch aus, dass die Strategien und Mittel von gestern nicht mehr genügen, die Herausforderungen von heute zu lösen, vor allem in Bezug auf langfristige Planung und Zielorientierung. Während früher eine durch das Finanzmanagement erfasste Jahresplanung mit dem Vergleich von Soll- und Ist-Zahlen Transparenz schaffte, können heute kurzfristige Ereignisse die ganze Planung auf den Kopf stellen. Manager und Führungskräfte müssen vor der Flut an Informationen und Anforderungen kapitulieren, weil diese keine prognostische Aussagekraft mehr besitzen. Agiles Management hilft in dieser Situation, einzelne Führungskräfte zu entlasten.

Was Unternehmen und Organisationen in Zukunft benötigen, sind Mitarbeitende, die weniger und kurzfristiger planen, offen denken, vielfältiges Expertenwissen verschiedener Perspektiven abholen, durch gemeinsames Experimentieren und Fehler rasch lernen und operative und strategische Exzellenz anstreben. Der Umgang mit der VUCA-Welt ist sehr anspruchsvoll und betrifft in erster Linie die strategische Stufe der Unternehmen.

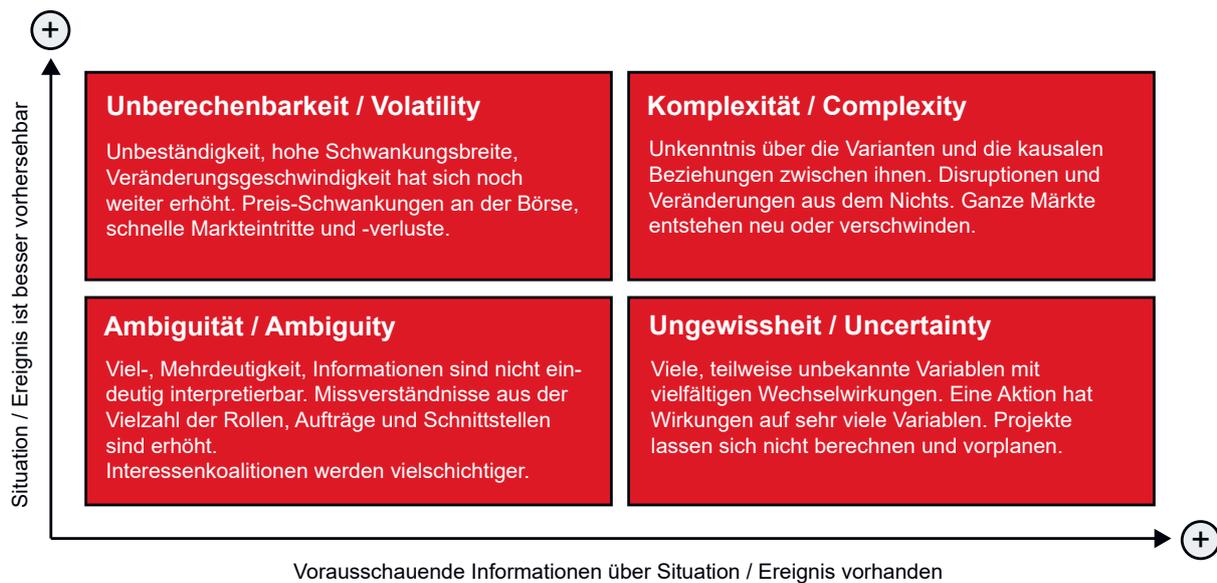


Abbildung 23: VUCA-Welt (Zäch, 2021)

Trotzdem sind die Gedankenspiele, welche sich mit dem Umgang der VUCA-Welt verbinden, eng mit dem Mindset der Agilität verwoben.

Agile Führung

Agile Führung bedeutet, immer daran zu arbeiten, dass sich ein Unternehmen an Veränderungen maximal schnell anpassen kann. Sie unterstützt Mitarbeitende auf Augenhöhe dabei, gemeinsam die besten Lösungen für Herausforderungen zu finden. Dabei finden die Bedürfnisse aller Stakeholder Berücksichtigung. Zentrale Werte der agilen Führung sind Offenheit für Neues, Kommunikation und flache Hierarchien.

Die wichtigsten Prinzipien der agilen Führung sind folgende:

1. Offenheit gegenüber Veränderungen
2. Ganzheitliche Betrachtungsweise
3. Dem Team vertrauen und Arbeit abgeben
4. Entscheidungen werden agil (schnell) gefällt
5. Iterationen können sinnvoll sein
6. Etablierung einer starken Feedbackkultur
7. Positiver Umgang mit Fehlern (offene Fehlerkultur)
8. Förderung von Kontakt und Kommunikation im Team
9. Änderung des Hierarchieverständnisses (hierarchische Führung vs. flache Hierarchie)
10. Anwendung agiler Methoden

Einführung agiler Werte

Neben der Vielzahl an Kompetenzen geht es im Rahmen der digitalen Transformation auch um die Entwicklung agiler Werte.

Diese Werte lassen sich in drei Stufen unterscheiden. Zu beachten ist, dass mit der Einführung der agilen Führung mindestens die mittlere Stufe erreicht sein muss. Ist dies nicht der Fall, sind diese Stufen vorab zu entwickeln. Vorab ist also die aktuelle Stufe der Organisation auf Basis der obigen Darstellung zu ermitteln und abzubilden. Dabei wird rasch sichtbar, dass im Unternehmen unterschiedliche Stufen vorhanden sind. Dies macht den angestrebten Kulturwandel anspruchsvoll, weil kein gemeinsames Fundament vorhanden ist.



Abbildung 24: Elemente agiler Werte (Zäch, 2021)

Das formale Fundament agiler Führung ist der Sinn und das gemeinsame Ziel der Arbeit. Vielfach wird dabei nicht berücksichtigt, dass alle Beteiligten eine Vertrauensbasis auf Sach- und Sozialebene aufbauen müssen. Vorbehalte und Bedenken müssen ausgeräumt werden, damit eine vertrauensvolle Zusammenarbeit entstehen kann.

Bei der Vertrauensbildung ist es wichtig, die Mitarbeitenden bei folgenden Kompetenzen zu unterstützen:

- Offenheit fördern, z. B. durch informellen Austausch, einfachere Projekte und Wissensaustausch
- Konfliktbereitschaft, z. B. durch die Definition von Regeln, geschützten Räumen/Rückzugsmöglichkeiten
- Selbstverpflichtung und Kommitment, z. B. durch gemeinsame und kommunizierte persönliche Zielsetzungen/Unterscheidung von Handlungs- und Verhaltenszielen
- Gegenseitige Verpflichtung, z. B. durch kommunizierte Verbindlichkeit
- Zielorientierung, z. B. durch Rückzugsorte zum konzentrierten Arbeiten

Vertrauensvolle Beziehungen in Teams zeichnen sich dadurch aus, dass sich Teammitglieder sehr gut kennen, Persönliches kennen und auch bei Krisen die persönlichen Kenntnisse vertraulich behandeln. Die beschriebenen Veränderungen können nur evolutionär bewerkstelligt und auf keinen Fall angeordnet werden. Vertrauen aus Sicht der Führung zeigt sich im Vertrauensvorschuss, in der Delegation von Kompetenzen, Verantwortung und der Wertschätzung für geleistete Arbeit.

13.2 Change-Management

«Die einzige Konstante im Leben ist die Veränderung.» (Heraklit, 600 v. Chr.)

Das Zitat des griechischen Philosophen Heraklit ist heute aktueller denn je. Es bringt auf den Punkt, dass der Wandel unausweichlich ist und die Menschen lernen müssen, ständige Veränderungen zu akzeptieren.

Wie bereits erwähnt, führt die Digitalisierung zu einer weltweiten Veränderung in einer nie dagewesenen Geschwindigkeit. Weder die Unternehmensgrösse, die Marke, die Tradition noch die Erfahrung können Unternehmen vor der Digitalisierung schützen. Dieses unberechenbare und schnelllebige Umfeld wird auch als VUCA-Umfeld bezeichnet. VUCA steht für ein volatiles, unsicheres, komplexes und mehrdeutiges Umfeld.

Der Wandel selbst ist ein komplexes Phänomen, das durch den Gebrauch von Methodiken wie dem Change-Management vereinfacht werden soll. Aufgrund der Komplexität, der Geschwindigkeit und der Tiefe der Veränderungen greift aber im Rahmen einer digitalen Transformation das bewährte Change-Management zu kurz. Der Grund dafür ist, dass die «alten» Change-Ansätze in vielen Unternehmen als Prozess gelebt, der Top-down vorgegeben und umgesetzt wird. Ausserdem stammt das klassische Change-Management aus einer Zeit, in der die technologischen Brüche nicht so stark waren wie heute.

Change-Management kann also nicht mehr als eine momentane Aktion angesehen werden, da sich alles schneller ändert. Die Welt fängt gerade erst an, die technologischen Möglichkeiten zu entdecken und zu erleben. Hinzu kommt, dass sich die Welt von einer analogen zu einer digitalen Umgebung wandelt. Die neuen Technologien bieten eine Vielzahl von digitalen Mitteln, mit denen typische Probleme im Change-Management, etwa Kommunikationsprobleme, in Zukunft vermieden werden können.

Im Rahmen der digitalen Transformation reicht es heute nicht mehr aus, ein einfaches Prozess-Redesign oder eine Technisierung zu adressieren (Capgemini-Consulting, 2017). Es braucht daneben als wesentliche Voraussetzung auch den Einbezug der Unternehmenskultur und die Beleuchtung dessen, was die optimale Kultur für die digitale Welt der entsprechenden Unternehmung ist. Zudem wurde festgestellt, dass es in der Regel an Kommunikation mit den Mitarbeitenden in der Transformationsphase mangelt und damit zusammenhängend der Umgang mit Ängsten der Mitarbeitenden unzureichend ist. In diesem Zusammenhang hat sich auch unter dem agilen Change-Management das **8-Stufen-Modell** (Kotter, 1996) bisher sehr bewährt. Dabei gilt zu

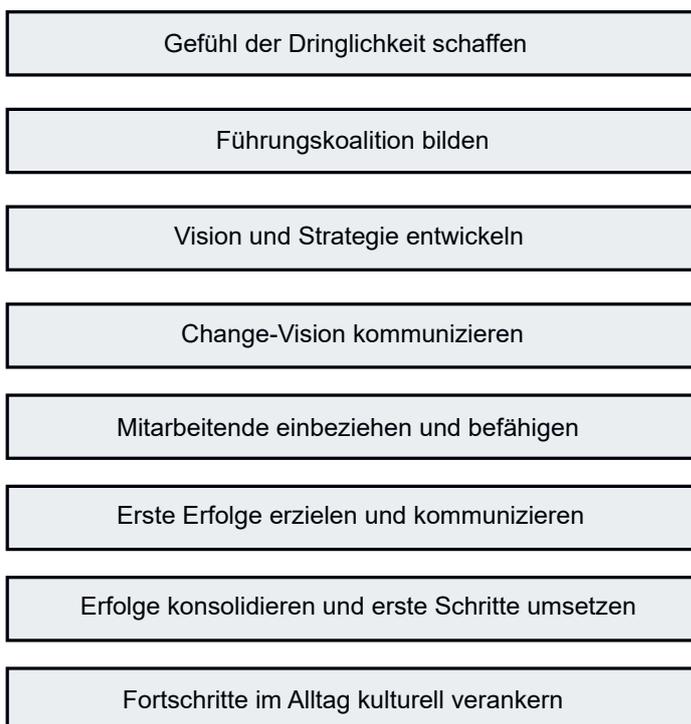


Abbildung 25: 8-Schritte-Modell (Kotter, 1996)

beachten, dass die 8 Schritte bereits früh in den Change-Prozess einzubinden sind, um die sich laufend ändernden Anforderungen des dynamischen Umfelds besser in Einklang zu bringen.

Die Zeiten von umfangreichen und trägen Change-Initiativen gehören im digitalen Zeitalter der Vergangenheit an. Analog der Kollaboration in den Kernprozessen eines Unternehmens muss auch das Change-Management agiler und somit zu einer Selbstverständlichkeit werden.

Agiles Change-Management sieht so aus, dass in kurzen Abständen, fokussiert auf den gewünschten Erfolg, definierte Veränderungsschritte realisiert werden. Veränderungen werden den Mitarbeitenden gegenüber laufend kommuniziert. Die Aufteilung der Veränderung erfolgt in kleinen Schritten, um diese für die Mitarbeitenden fassbar zu machen und ihnen die Möglichkeit zu geben, Sicherheit und Vertrauen zu haben. Kontinuierliches Stakeholder-Management ist dabei zentral. Review-Momente am Ende jedes Veränderungsschrittes helfen, den Benefit der Umsetzung fortlaufend herauszustreichen, und bieten zudem die Möglichkeit, Justierungen kurzfristig vorzunehmen. Ausserdem kann so der folgerichtige, für die Unternehmung verdaubare Veränderungsschritt identifiziert und entsprechend geplant werden.

14 Literaturverzeichnis

Capgemini-Consulting. (2017).

ExpertFocus. (2019/12). EXPERTsuisse.

Gartner. (27. Juli 2020). Gartner Research. Von

<https://www.gartner.com/en/documents/3988021/magic-quadrant-for-robotic-process-automation>
abgerufen

Kotter, P. (1996). 8-Stufen-Modell.

Martin, A. (2008). Digital Literacy for the Third Age: Sustaining Identity in an Uncertain World.

Microsoft Business Solutions. (2005). ERD Microsoft NAV.

O'Brien, J. A. (2010). Management Information Systems.

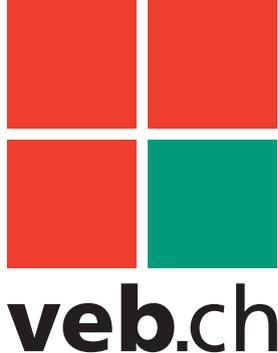
Ruoss, S. (2015). Von <https://svenruoss.ch> abgerufen

[tableau.com](https://www.tableau.com/de-de/learn/articles/was-ist-big-data). (2021). Von <https://www.tableau.com/de-de/learn/articles/was-ist-big-data> abgerufen

Zäch, M. (2021). Projektmanagement in ERP-Systemen. St. Gallen, Schweiz.

15 Abbildungsverzeichnis

Abbildung 1: Die Architektur von Cubetools	10
Abbildung 2: Beispiel OLAP-Cube	10
Abbildung 3: Vorsystem und Data Warehouse.....	13
Abbildung 4: Cubetool – Data Warehouse	18
Abbildung 5: Beispiel systematischer Aufbau eines Data Warehouse	19
Abbildung 6: Prozess-Sicht eines ERP-Systems.....	20
Abbildung 7: ERP-Organisation	21
Abbildung 8: ERD-Diagramm (Microsoft Business Solutions, 2005)	25
Abbildung 9: Software-Verarbeitungsebenen (Zäch, 2021)	27
Abbildung 10: Aufbau Debitorenbuchungen (Zäch, 2021)	28
Abbildung 11: Entscheidungsebene und Information (O’Brien, 2010)	30
Abbildung 12: BI back/Frontend	31
Abbildung 13: Natur einer Künstlichen Intelligenz.....	50
Abbildung 14: Differenz Supervised Learning/Unsupervised Learning	52
Abbildung 15: Schritte von Daten bis zum Testing	54
Abbildung 16: Decision Tree.....	57
Abbildung 17: Confusion Matrix	58
Abbildung 18: Illustration eines Explainable AI für einen Fall	62
Abbildung 19: Digitale Transformation als Teil der Digitalisierung	64
Abbildung 20: Levels of Digital Literacy (Martin, 2008).....	65
Abbildung 21: Ziele der digitalen Transformation (Ruoss, 2015)	66
Abbildung 22: Digital Finance Transformation	68
Abbildung 23: VUCA-Welt (Zäch, 2021).....	72
Abbildung 24: Elemente agiler Werte (Zäch, 2021)	73
Abbildung 25: 8-Schritte-Modell (Kotter, 1996).....	74



veb.ch – Fachwissen und Erfahrung aus erster Hand

veb.ch ist Ihre erste Anlaufstelle für alle Fragen in den Bereichen Rechnungslegung, Controlling und Rechnungswesen.

Unser Verband umfasst schweizweit über 9'000 Mitglieder. Dazu zählen insbesondere Expertinnen und Experten in Rechnungslegung und Controlling sowie Inhaberinnen und Inhaber des Fachausweises im Finanz- und Rechnungswesen. Sie zählen in der Schweiz zu den staatlich geprüften Fachleuten für alle Fragen des Rechnungswesens auf allen Unternehmensebenen. Daneben können auch ausgewiesene Spezialistinnen und Spezialisten des Rechnungswesens mit anderer Ausbildung unserem Verband beitreten. Als Passivmitglied sind alle mit uns verbundenen Personen bei veb.ch herzlich willkommen.

Bildung liegt uns besonders am Herzen. Deshalb sorgen wir als Verband laufend dafür, dass Ausbildungen und Prüfungen den aktuellen Bedürfnissen angepasst werden. Zusammen mit dem Kaufmännischen Verband führen wir die jährlichen Prüfungen für Fachleute im Finanz- und Rechnungswesen durch.

Zu unserem Weiterbildungsangebot:

 [www.veb.ch / Seminare und Lehrgänge](http://www.veb.ch/Seminare_und_Lehrgaenge)

veb.ch

Schweizerischer Verband der dipl. Experten in Rechnungslegung und Controlling und der Inhaber des eidg. Fachausweises in Finanz- und Rechnungswesen. Seit 1936

Talacker 34
8001 Zürich
Telefon 043 336 50 30
info@veb.ch

Lesen Sie unseren Blog unter
blog.veb.ch

Besuchen Sie unsere digitale Welt auf
www.veb.digital

Folgen Sie uns auf:



EDU  **UA**

Schweizerisches Qualitätszertifikat für Weiterbildungsinstitutionen
Certificat suisse de qualité pour les institutions de formation continue
Certificato svizzero di qualità per istituzioni di formazione continua