



# Business Analytics I

## Hintergrund und Konzept der BA Professionalisierung

Dr. Holger Steinmetz  
Lehrstuhl für Unternehmensführung  
Universität Trier

- Umwelt des Mittelstands von 1950 - 1980
  - **Fokus auf nationale / regionale Märkte** und weniger internationale Beziehungen und Geschäfte
  - **Weniger (internationale) Konkurrenz**
  - **Höhere Regulierung** v.a. in europäischen Ländern → Klare Regeln, Vorgaben und Planbarkeit
  - **Industrieller Wandel:** Verschiebung von traditionellen Wirtschaftssektoren hin zu neuen Technologien und Industrien → Chancen für Wachstum
  - **Traditionelle Unternehmensstrukturen und -kultur:**
    - **Struktur:** Formalisierung, Zentralisierung von Kontrolle
    - **Kultur:** Performanceorientierung, geringe Innovationskultur, auf langfristige Planung und Sicherheit fokussiert ("internal process culture")

- **Mehr Wettbewerb** durch Globalisierung und Internet
- **Geringere Regulierung** (Flexibilität)
- **Veränderte Kundenanforderungen**
  - Kunden sind informierter über Konkurrenzprodukte und Preise
  - Kunden haben heute **höhere Anforderungen** an Qualität, Service und Personalisierung.
  - **Veränderte Werte** z.B. Nachhaltigkeit
  - Höhere switching-Wahrscheinlichkeit, geringere Kundenloyalität
- **Digitalisierung**
  - Neue **Geschäftsmodelle** und mehr Möglichkeiten der **Vermarktung** (z.B. social media)
  - Neue **Verpflichtungen** (→ Kunden erwarten digitale Optionen)

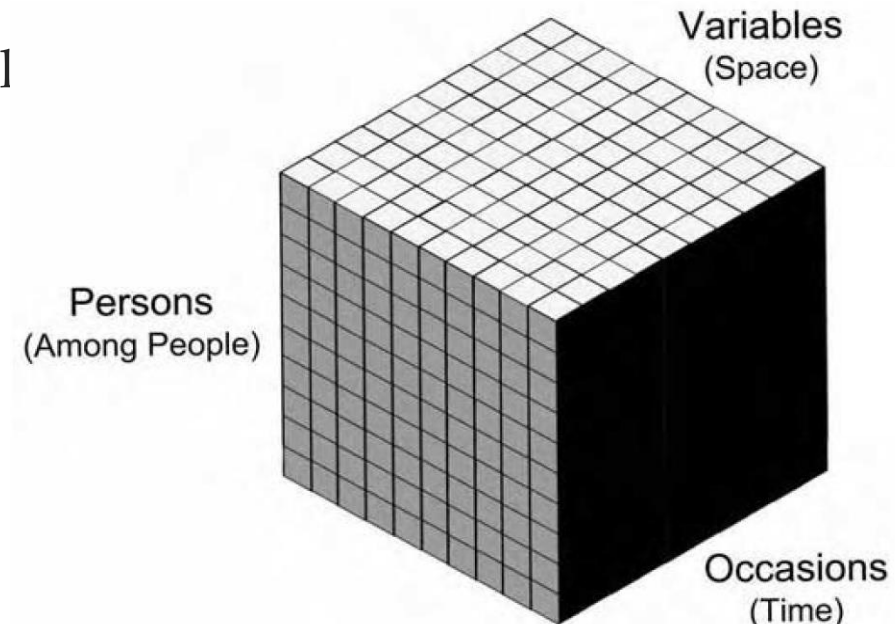
→ Was genau ist DIE Umwelt?

- Enthält den Teil der Realität, der für die **eigene Handlungsfähigkeit** relevant ist—d.h.
  - die **Wahl** (→ Ziele, Entscheidungen)
  - **Art/Ausformung** (→ Handlungsausführung, Adaptation an Veränderungen, Effizienz)
  - **Erfolg** (→ Barrieren, Gefahren, Erleichterungsfaktoren)

→ Handlungspsychologie: "**Handlungsfeld**"

- Verschiedene Modelle und Taxonomien. Zentral:
  - **Complexity**: Anzahl der relevanten Aspekte der Umwelt (z.B. Gesetze, Kunden-Diversität)
  - **Munificience**: Ressourcenreichtum vs. Sättigung
  - **Uncertainty**: Ungewissheit über die
    - **Zukunft** und
    - **Handlungs-Ergebnis-Kontingenzen**
  - **Hostility**: Gefährdet die Existenz der Organisation (z.B. intensiver Wettbewerb, harsche Kunden)
- Gibt es sowohl **extern** als auch **intern**

- Big Data: Die 3 Vs
  - **Volume:** Menge an Daten
  - **Variety:**
    - Datenquellen, z.B. Kunden (digital trace data), Maschinen
    - Vielfalt z.B. durch Zahlen, Text, Bilder, Audio
  - **Velocity:** z.B. durch Sensoren, automatisch/digital Prozesse
- Volume: **Cattel's Data Cube**
  - Large N (Fälle, z.B. Kunden)
  - Large P (Variables/Features)
  - Large T (Zeitpunkte, z.B. Sensordaten)



- **Definition BA:** *"Delivering the right decision support to the right people at the right time"*

→ BA unterstützt **Entscheidungs- und Handlungsprozesse**

- **Beispiele:**

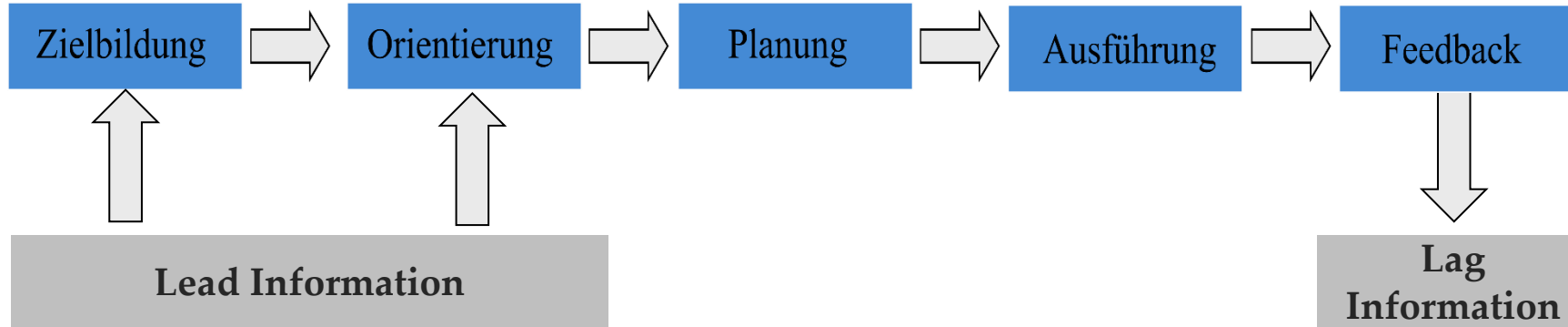
- Welche Leute sollten wir bei der Besetzung einer Stelle auswählen?
- Was ist der Erfolg einer Marketingkampagne?
- Welche Bedürfnisse/Sichtweisen haben unsere Kunden?
- Wie können wir Ressourcen sparen?
- Wie können wir Ausfälle von Maschinen vermeiden?

- **Exkurs Handlungsregulationstheorie:** Handlungen sind

- **Sequentiell:** Vollziehen sich über Phasen über die Zeit
- **Hierarchisch:** Organisiert über Zielhierarchien

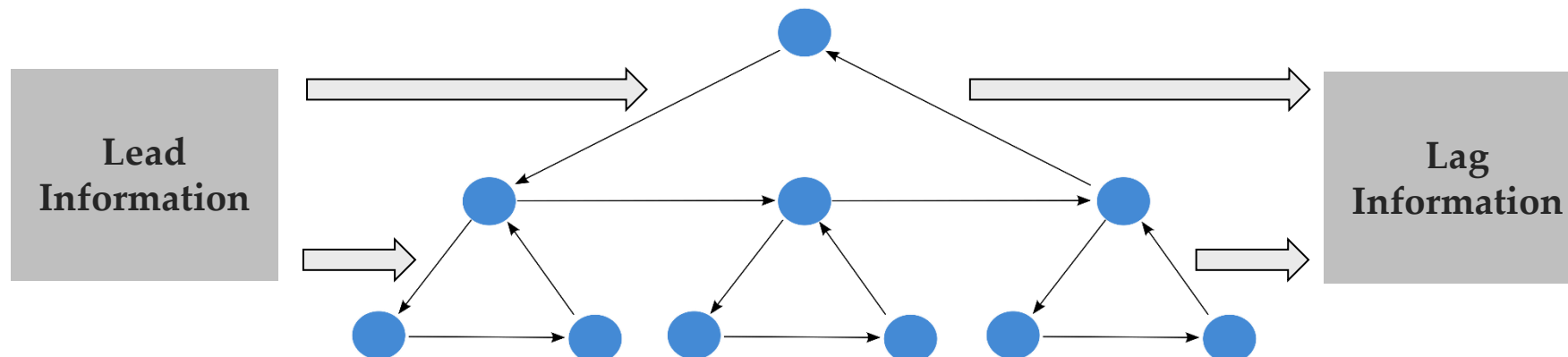
→ BA zielt darauf ab, an entscheidenden Punkten einer Handlung nötige Information zu liefern | 6

- Handlungssequenz



- Hierarchische Zielorganisation:

- Das allgemeine Ziel (s.o.) wird **operative Subziele** (~ to do's) zerlegt
- Der Handlungsprozess läuft über **feedback loops**, bis das (Sub)Ziel erreicht ist.



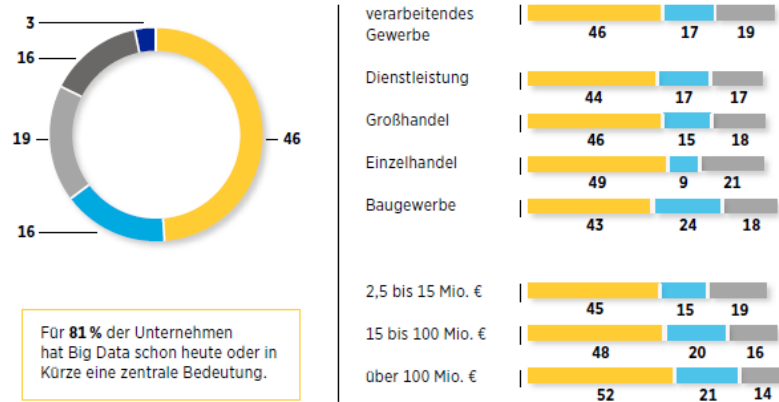


# BA Professionalisierung



- Laursen & Thorlund: "The BA model"
  - Beschreibt die ideale Integration von BA in eine Organisation
  - Realität ist sehr weit davon entfernt (→ **BA Reifegrad / maturity model**)
  - **Evidence-management Konzept** in der angewandten Forschung:
    - Manager lesen keine wissenschaftliche paper
    - Orientieren sich eher an Moden und Mimikri der Konkurrenz
    - Daten bedrohen die eigenen Position / Macht
    - Effektivitätseinschätzung auf Basis von Plausibilität
    - Im Fall speziell des Mittelstandes: Kulturelle / historische Aspekte

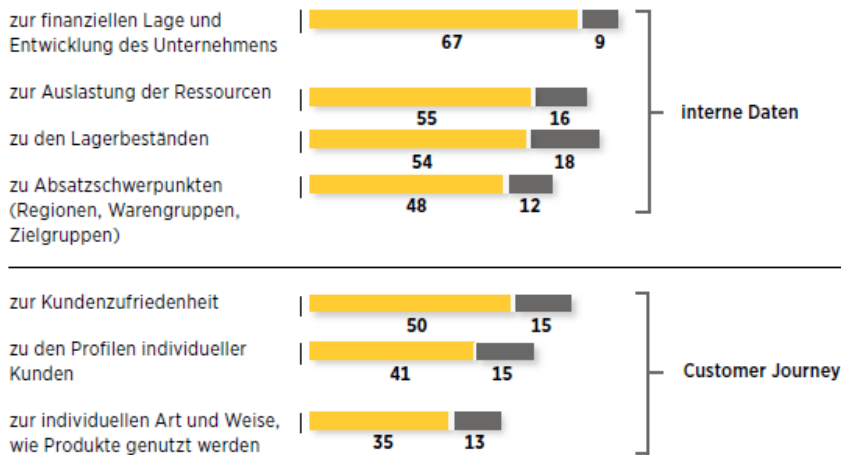
## Welche Bedeutung hat die generelle Zunahme digitaler Daten für die Unternehmen?



■ schon heute zentral ■ in Kürze zentral ■ in Zukunft zentral ■ noch nicht abzuschätzen ■ für uns nicht relevant

## Welche digitalen Daten liegen den Unternehmen vor?

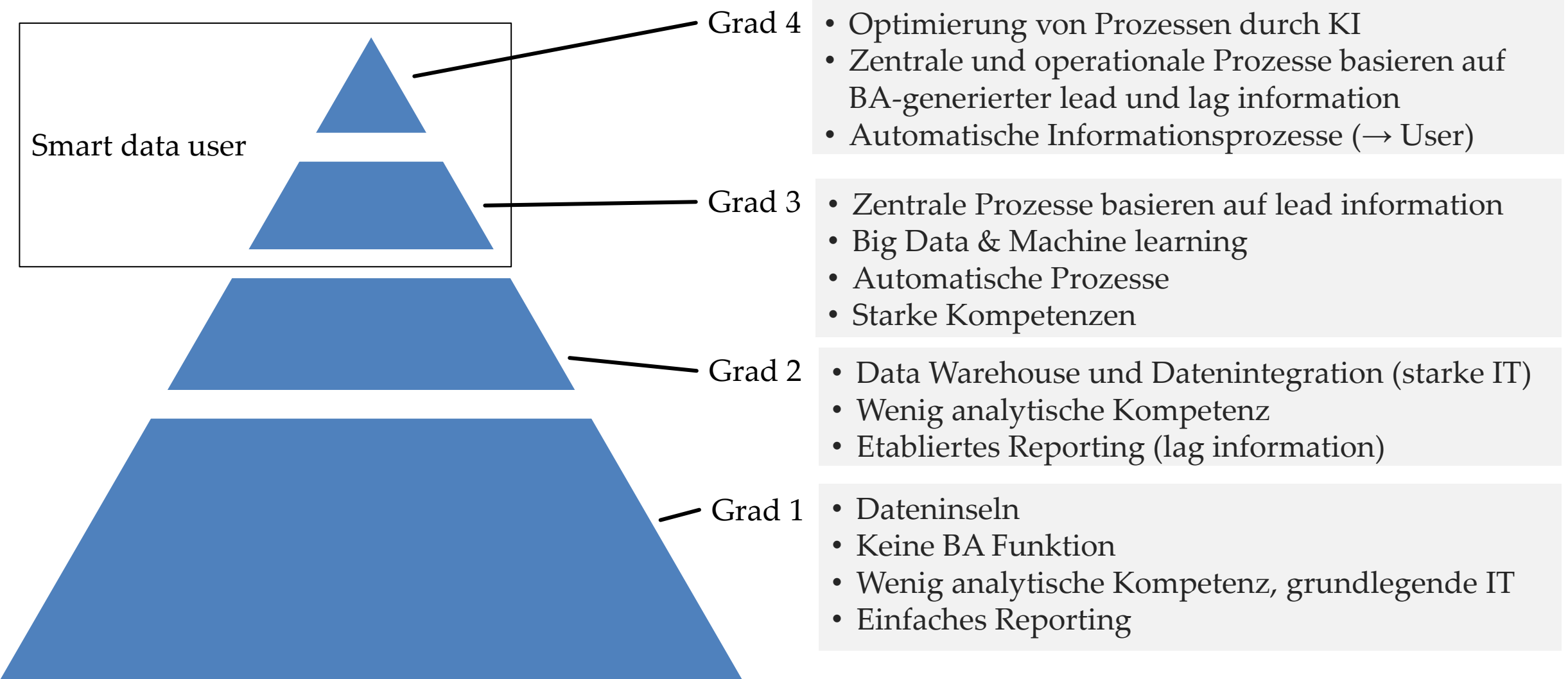
### Die (weitere) Erfassung von Daten ...



■ findet statt ■ ist in Planung

## • Commerzbank-Umfrage

- 81% finden Big Data zentral (jetzt oder zukünftig)
- Für alle Branchen und Größen relevant
- Aber (nicht in der Abb.): Nur 8% analysieren Daten systematisch ("smart-data user")
- Untere Abbildung: Am Nicht-Vorhandensein digitaler Daten liegt das nicht:
  - 67% erheben digitale Daten über finanzielle Lage
  - ~50% über Ressourcenauslastung, Lagerbestände und Zielgruppen
  - Nur 35-50% Analyse von Kundenprofilen, Produktverwendung und Zufriedenheit (!)



Competencies, people, and processes to create successful business intelligence and analytics

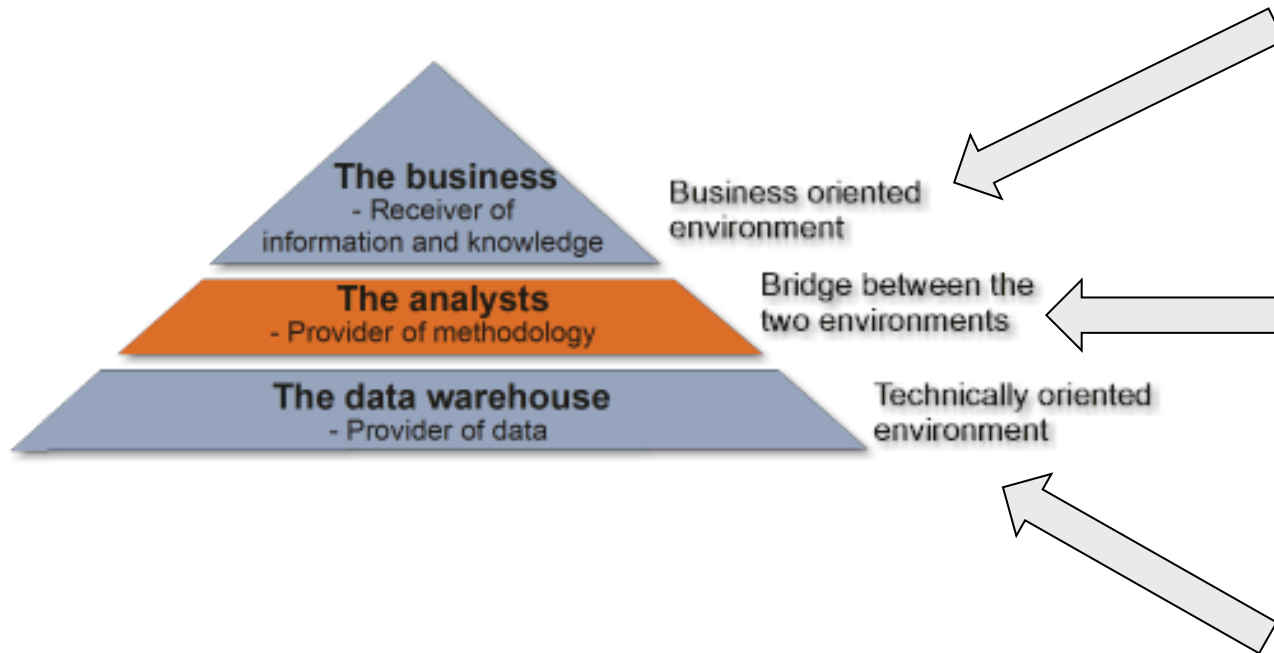
## Business-driven environment

(Top) Management <i>Chapter 2</i>	Strategy creation Choosing an information strategy
Operational decision makers <i>Chapter 3</i>	Business processes Using information and knowledge
Analysts, controllers, and report developers <i>Chapter 4</i>	Reporting and analytics Creating information and knowledge
ETL developers and database specialists <i>Chapter 5</i>	Data warehouse Gathering data, making them accessible and usable
IT professionals <i>Chapter 6</i>	Data sources and IT Infrastructure Data creation

Information supply

## Technically oriented environment

- Zeigt
  - die **Rollen** und
  - **Aufgabenbereiche**die für BA relevant sind
- Oben: Businessorientierung
- Unten: Technik/Datenorientierung
- Von oben nach unten: Informationsbedarfe
- Von unten nach oben: Informationsangebot



- **Businessorientierung**

- Top Management
- Funktionsbereiche (FB), z.B. HR, Marketing, Produktion, Finanzen
- Fokussiert auf **strategische** und **operative Prozesse**

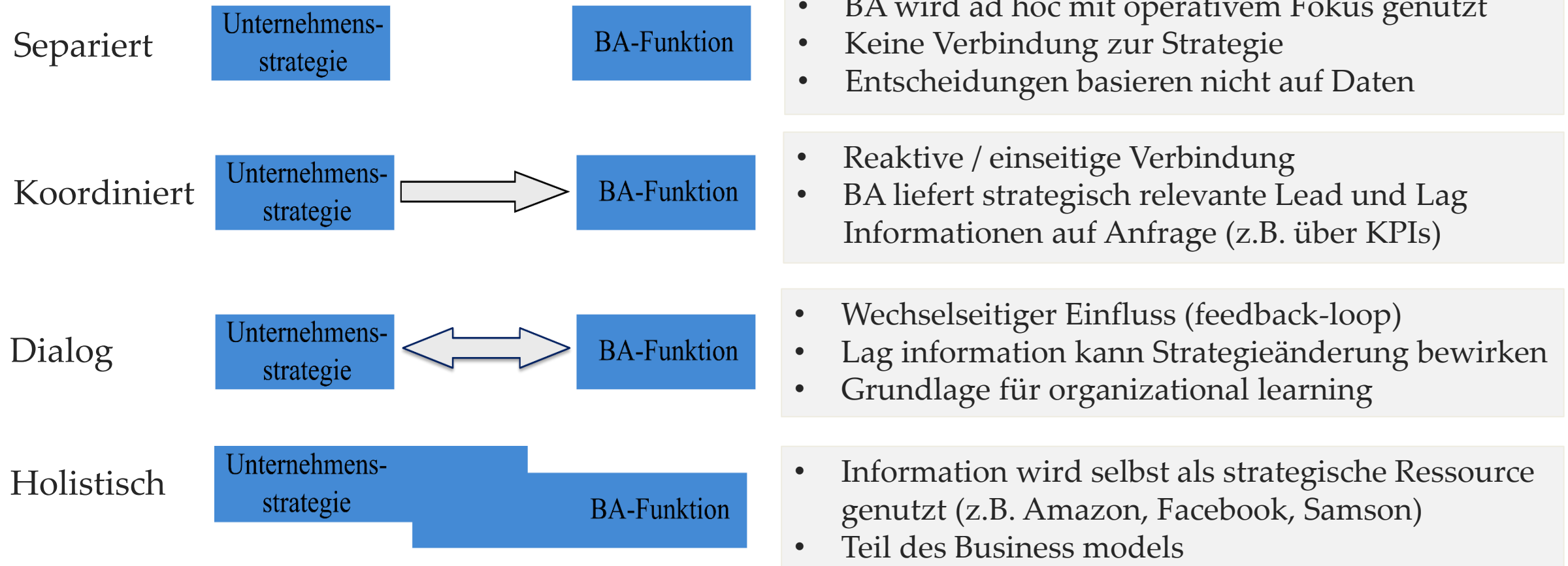
- **Analyseorientierung**

- Arbeitet mit Daten und liefert Information
- **Integriert** die Business-, Daten- und Analyseorientierung und **vermittelt**

- **Datenorientierung**

- Sammlung und Integration der Daten
- ETL (extract, transform, load): Cleaning, Integration und "wrangling"

- Betrifft die Frage der Wertschöpfung durch die BA-Funktion
- Dimension von "wenig" (BA-Information als "nice to have") bis zu "stark relevant" (BA-Information ist zentral für strategische Entscheidungen)
- Laursen & Thorlund: **4 Typen der Integration**



- **Informationsstrategie:** Wie kann man den geäußerten Informationsbedarf decken?

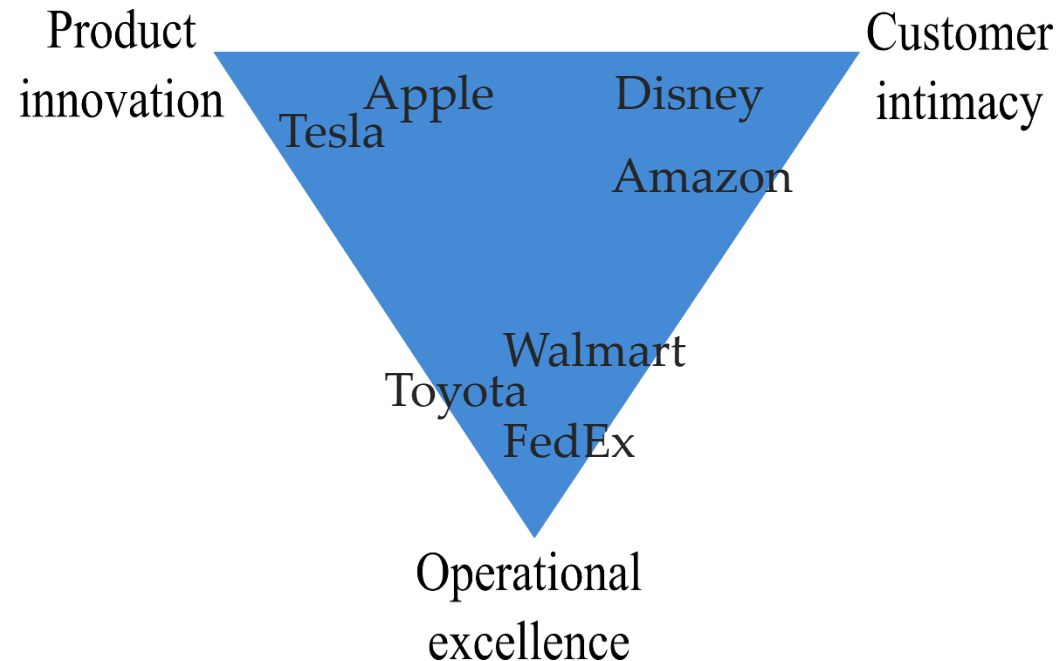
→ Spezifikation der 4 zentralen Zutaten eines BA-Prozesses:

- (1) **Ziel:** Was möchte man wissen? → Reflektiert den Informationsbedarf (z.B. etwas über Kunden zu lernen)
- (2) **Inhalt:** Welche Phänomene sind relevant (z.B. Kundenzufriedenheit)?
- (3) **Zugangsweg:** Welche Daten kommen in Frage (z.B. Survey vs. Social Media)?
- (4) **Analytisches Design:** Was macht man mit den Daten um das Ziel zu erreichen? (z.B. Clusteranalyse, Zeitreihenanalyse etc.)
- (5) **Nutzung der Ergebnisse:** Ergebnisbericht? Präsentation? Dashboard?  
Fundament für einen automatischen KI-Prozess?

# Ziel: Was möchte man wissen?

- Informationsbedarfe variieren in der Wichtigkeit → Welche sind von hoher Wichtigkeit?
- Strategie-Taxonomie von Treacy & Wiersema (1993)

Treacy, M., & Wiersema, F. (1993). Customer intimacy and other value disciplines. *Harvard Business Review*, 71(1), 84-93.  
(<https://hbr.org/1993/01/customer-intimacy-and-other-value-disciplines>)



- Nicht alle Dimensionen können maximiert werden (wobei das Dreieck Konflikte überbetont)
- Je nach Orientierung werden Informationen verschiedene Prioritäten/Werte haben
- Das eine Organisation einen Fokus hat, bedeutet nicht, dass andere Dinge unwichtig sind (sie sind nur nicht *der* zentrale Aspekt)



- Die Taxonomie von Treacy und Wiersema kann sehr gut verwendet werden, um Informationsbedarfe (→ **Ziele der Analyse**) zu ordnen und zu verstehen
  - **Produktorientierung:** Analysen dienen dazu, Produkte zu verbessern oder neue Ideen zu generieren
  - **Kundenorientierung:** Analysen dienen dazu, etwas über Kunden zu lernen
    - Wie ticken Kunden?
    - Welche Bedürfnisse, Werte, Interessen haben sie?
    - Welche (finanziellen) Ressourcen haben sie?
  - **Prozessorientierung:** Analysen dienen dazu, Prozesse zu optimieren:
    - Fehler reduzieren
    - Effizienz erhöhen

- Beispiele

- **Entwicklung des Marktes und Verkaufszahlen** (z.B. mittels Zeitreihen).
- **Analysen von Unzulänglichkeiten oder Beschwerden** über Funktionsaspekte des Produkts etc. (z.B. auf Social Media, Beschwerdhotlines, Rezensionen) und NLP-Methoden.
- **Identifikation neuer Märkte** über Identifikation von Kundensegmenten und Prognose ihrer Entwicklung mittels Clusteranalyse.
- **Identifikation von Synchronitäten vs. Substitute** von Produkten oder Teilen, die mit dem Produkt assoziiert sind (mittels Market Basket-Analyse).
- **Industry 4.0 und Internet of Things (IoT)**: Produktgebrauch oder Probleme (z.B. crashreports, Analyse des Klickverhaltens/Cookies auf Webseiten) mittels Zeitreihen, machine learning etc.

- Ziel ist guter Service (Jeff Bezos: "Customer Obsession") damit Kundenzufriedenheit und Loyalität
- V.a. essentiell, wenn keine Differenzierung über das Produkt möglich ist (Banken, Versicherungen etc.)
- **Beispiele über Kunden-Informationen** (→ Inhaltsdimension in der Informationsstrategie)
  - **Demografie** (Alter, Geschlecht, Familienstand, Bildung, Beruf, Standort usw.)
  - **Psychografie** (Werte, Einstellungen, Interessen, Hobbys, Lebensstil, Persönlichkeit usw.)
  - **Verhaltensdaten** (Kaufhistorie, Häufigkeit, Betrag, Markenloyalität, Kaufmuster, Kanalpräferenzen usw.) → Wächst v.a. im Rahmen der Digitalisierung ("digital trace data")
  - **Zufriedenheit**
  - **Net Promoter Score (NPS)**: Wie wahrscheinlich ist es, dass ein Kunde das Unternehmen anderen empfiehlt?

- **Analyseformen:**
  - **Trends von Präferenzen** mittels Zeitreihen
  - **Vorhersage des Customer life time values (CLV)** mittels Machine Learning auf Basis von Kundeninformationen auf Basis der Demographie, Psychographie oder Verhalten)
  - **Market segmentation** mittels Clusteranalyse
    - **Need-based:** Cluster von Kunden mit bestimmten Präferenzen
    - **Value-based:** Cluster mit "wichtigeren" vs. "weniger wichtigen" Kunden (→ CLV)
  - **Churn prediction:** Vorhersage der Dauer, bis der Kunde kündigt und der Ursachen mittels Survivalanalyse.

- Ziele
  - Effizienzsteigerung — d.h. Ressourcen zu schonen
  - Stabilität erhöhen (z.B. Schwankungen und Ausfälle vermeiden), vgl. Umweltdimension "Unsicherheit"
  - Effektivität/Qualität steigern (Fehlerrate senken)
  - Langfristig:
    - Preise senken können
    - Kundenzufriedenheit erhöhen (z.B. über Qualität, Preise, Lieferzeit, Möglichkeit der Retouren)

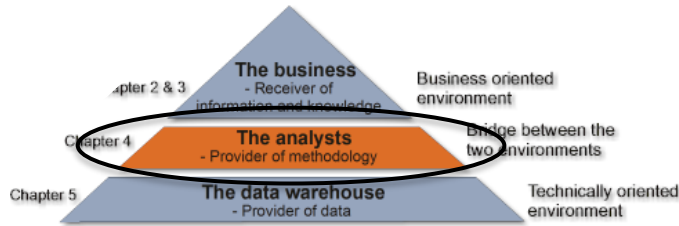
- **Ansätze im technischen Bereich**
  - **Analyse und Vorhersage von Bestellungen** (→ Lean Management) mittels Zeitreihen und Machine Learning (Beispiel: IntabPro: <https://www.intab.pro/>)
  - **Routenplanung in der Logistik** unter Berücksichtigung von Wetter, Verkehr, Uhrzeit
  - **Predictive Maintenance:** Basiert auf Sensordaten von Maschinen, z.B.
    - **Anomaly Detection:** Identifizieren außergewöhnlicher Spitzenwerte (oder Muster, Häufigkeiten) und automatische Benachrichtigung von Operateuren
    - **Survivalanalyse:** Beschreibung der Lebensdauer (hazard rate) und deren Vorhersage.
  - **Fraud detection:** Identifikation verdächtiger Transaktionen (→ Anomaly Detection), und deren Vorhersage (durch Verhaltens- und Kundendaten)

- **Ansätze im Humanbereich (Human Resource Management):**
  - Vorhersage der zu erwartenden Performance (im Recruiting-Kontext)
  - Automatisierung des Recruiting (Mögen Bewerber nicht!)
  - Analyse der Kompetenzen und Trainingsbedarfs (Wichtig bei Strategieänderungen, die Implikationen für das Kompetenzprofil der Belegschaft haben)
  - **Turnover Prediction** (vgl. Churn prediction). Wann kündigen welche Mitarbeiter und warum?
- **Nebenwirkungen / Probleme:**
  - Ethische Problematik (prediction error, systematischer bias/Diskriminierung → "AI fairness")
  - Datenschutzrechtliche Aspekte (DSGVO)
  - Negative "Interactional justice"-Wahrnehmungen (z.B. bei Bewerbern oder Mitarbeitern)



# Die Rolle der/des Analysten





- Der Analyst ist die **zentrale Brücke** zwischen Personen
  - mit einer **Business-Orientierung**—d.h.
    - **operativen und/oder strategischen** Zielen und Aufgaben
    - entsprechenden **Informationsbedarfen**
    - Perspektive der Organisation: System von "**value-adding processes**"
  - mit einer **Dateninfrastruktur-Orientierung**—d.h.
    - technischem Fokus
    - Sicherung der Effizienz, Verfügbarkeit und Sicherheit datenbezogener Prozesse
    - Perspektive der Organisation: **Technisches System von Informationsflüssen**
- Bei mangelnder Integration: Das Data Warehouse entwickelt ein Eigenleben (→ Daten ohne Nutzen, schlechte Usability)

- **Umgang mit Daten:** → Data wrangling, mit Datenbanken umgehen können
- **Methodenkompetenz:** Statistik, Modeling, Designs, kausale Inferenz, Visualisierung
- **Business-Kompetenz:** Unternehmen brauchen mehr als nur Statistik-Fachleute
- **Kulturelles Wissen:** Muss wissen, wie User ticken (d.h. deren Kenntnisse, Gewohnheiten, Präferenzen, Vorlieben, Vorurteile)
- **Kommunikationsfähigkeit:**
  - **Gegenüber den "Business-Leuten":** Fähigkeit, Ergebnisse und analytische Prozesse Usern zu erklären
  - **Gegenüber den "Daten-Leuten":** Informationsbedarfe erläutern

- Übersetzung eines (meist vagen) **Informationsbedarfes** in eine konkrete **Informationsstrategie**

→ Dialog mit dem entsprechenden Funktionsbereich

- **Vermittelt Konflikte** bzgl.
  - Sprache und Jargon
  - Langfristiger vs. kurzfristiger Orientierung
  - Geschwindigkeit vs. Akkuratheit
  - Fokus auf Daten vs. Fokus auf Informationsgehalt / Nützlichkeit

- **Informationsstrategie:** Wie kann man den geäußerten Informationsbedarf decken?

→ Spezifikation der 4 zentralen Zutaten eines BA-Prozesses:

- (1) **Ziel:** Was möchte man wissen? → Produkte, Kunden, Prozesse
- (2) **Inhalt:** Welche Phänomene sind relevant (z.B. Kundenzufriedenheit)?
- (3) **Zugangsweg:** Welche Daten kommen in Frage (z.B. Survey vs. Social Media)?
- (4) **Analytisches Design:** Was macht man mit den Daten um das Ziel zu erreichen? (z.B. Clusteranalyse, Zeitreihenanalyse etc.)
- (5) **NEU: Nutzung der Ergebnisse:** Ergebnisbericht? Präsentation? Dashboard?  
Fundament für einen automatischen KI-Prozess?

- **Betrifft die Daten**
  - **Zu berücksichtigende Aspekte**
    - **Verfügbarkeit:** Sind Daten vorhanden oder müssen sie generiert werden?
    - **Validität:** Betrifft Stärke des Zusammenhangs zwischen dem Phänomen / Konstrukt (vgl. 2. Aspekt der Informationsstrategie: Welche Phänomene sind relevant?). Z.B. bei Survey höher als bei Textmining von Beschwerdhotlines?
    - **Fehlerbehaftetheit der Daten ("Veracity")**→ Zeitlicher, finanzieller und personenbezogener Aufwand für das Data Cleaning
    - **Data wrangling / Feature Engeneering:** Aus Daten müssen (konzeptionell bedeutsame Variablen) generiert werden. Daumenregel: 80% des gesamten Zeitaufwandes.
- Kennzeichnet den **Konflikt zwischen Verfügbarkeit und Nützlichkeit** (Validität, Fehler, Ausmaß des Feature Engeneering)



# Die Rolle des Datawarehouses

- Daten werden traditionell in den Funktionsbereichen erzeugt und aufbewahrt
  - **Personalabteilung:** Personalakte, Bewerbungen, Abwesenheits- und Urlaubsdaten
  - **Einkauf:** Auftragsbücher
  - **Finanzen:** Rechnungen, Transaktionen
  - **Produktion:** Inventar, Produktionsdaten
- **Ungünstig**
  - Digitalisierung und Explosion der Datenmenge (Big Data)
  - Geringe Nützlichkeit (Ziel ist Archivierung)
    - "Dateninseln" / Keine Integration
    - Wenig Dokumentation ("Datenfriedhöfe")
    - Keine Aufbereitung (schlechte Qualität)

- Speicherung großer Datenmengen
- Entlastung der Funktionsbereiche
- Bessere Datenaufbereitung
  - Höhere Qualität
  - Transparenz durch Dokumentation ("Meta-Daten")
  - Verfügbarkeit
- Integration von Daten aus verschiedenen Quellen / Funktionsbereichen
  - Keine isolierten Dateninseln
  - Bessere Informationsausbeutung (Kombination macht Daten wertvoller)



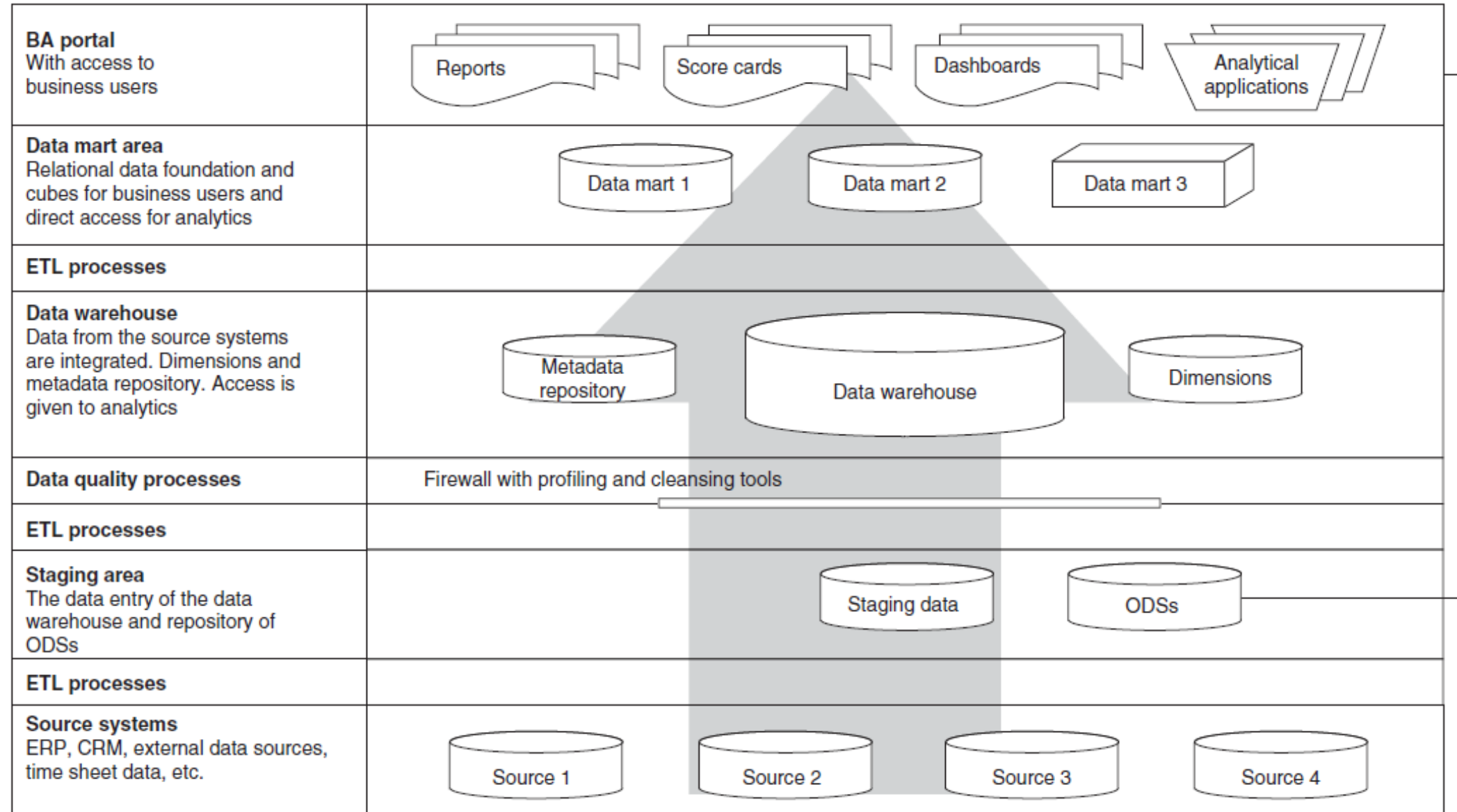
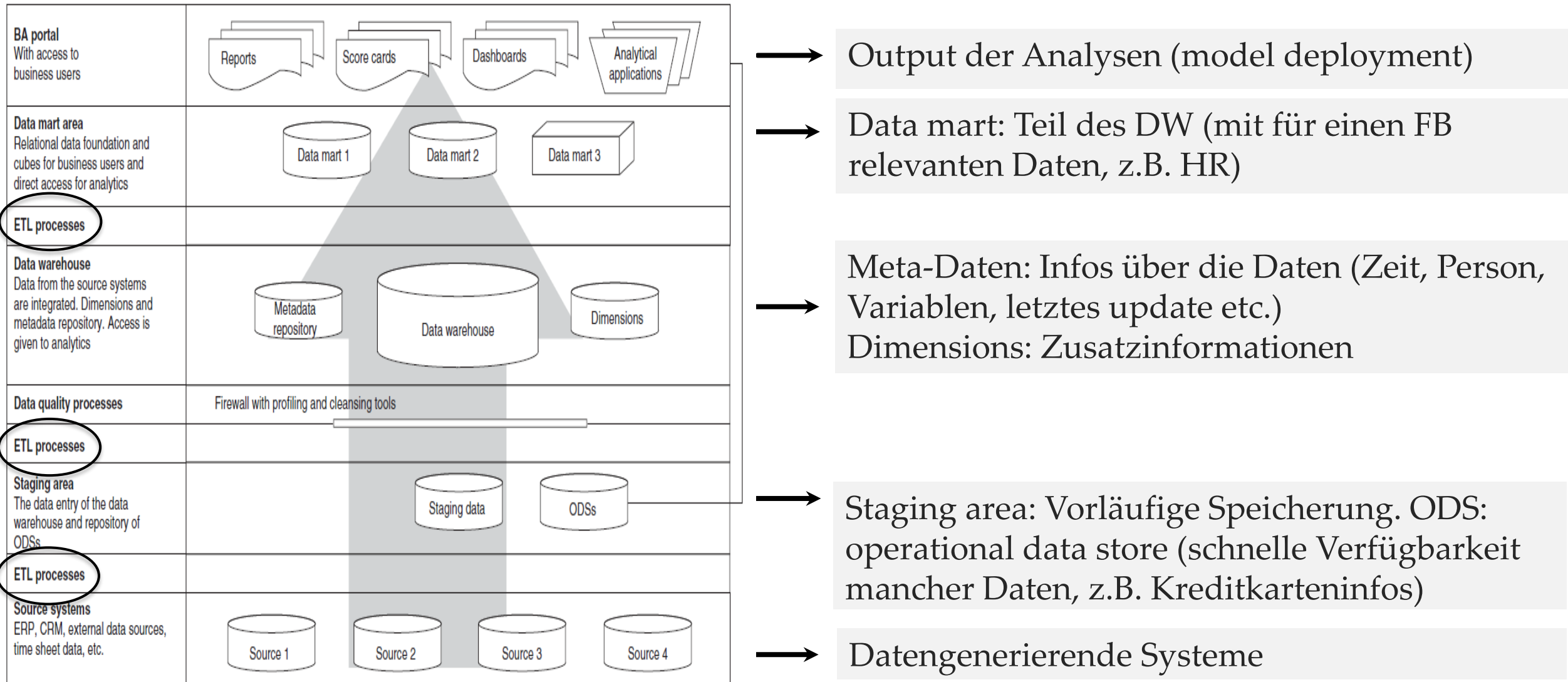


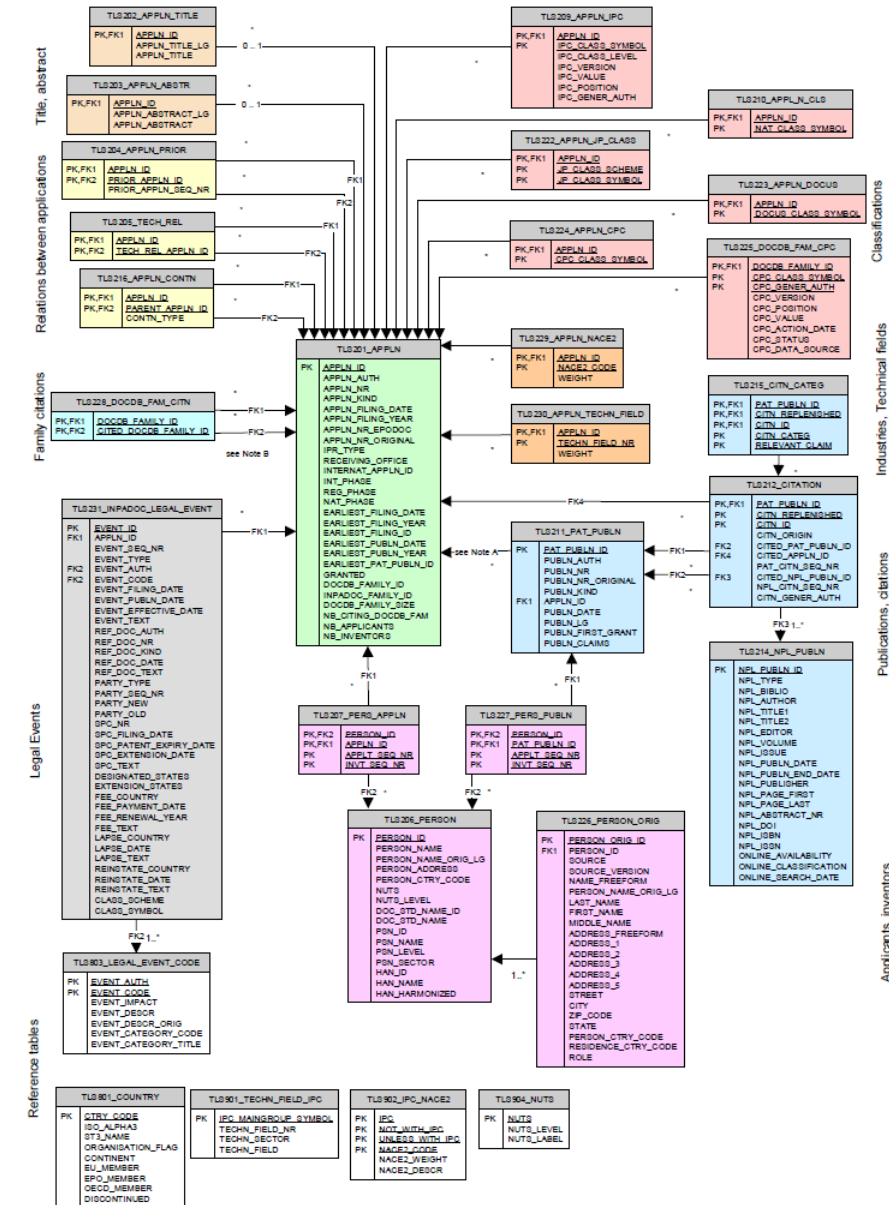
Exhibit 5.1 Architecture and Processes in a Data Warehouse

Datenquelle	Nutzung
<b>Transaktionsdaten</b> (z.B. Buchungen)	Analyse von Verkaufszahlen über die Zeit, automatisches Inkasso-System
<b>Sensordaten</b>	Produktverwendung, Predictive Maintenance
<b>Kundendaten</b> (Demos, Präferenzen, Interaktionen, z.B. Hotlines, Mails)	Segmentierung, Churn Prediction, Alarmsystem
<b>Mitarbeiter</b> (Rolle, Gehalt, Urlaub, Performance), Trainingsinformationen, Umfragen	Trainingsbedarf, Verhaltensvorhersage (z.B. Urlaub, Kündigung, Job Performance)
<b>Geo-Daten</b>	Benachrichtigungen, Fraud Detection, Logistik
<b>Social media</b>	Stimmung, Präferenzen, Trends, CSR, Tracking Bewerber
<b>Informationen über Maßnahmen</b> (z.B. Werbekampagnen)	Wissen was funktioniert (und für welche Art von Kunden)
<b>Web logs</b>	Kundenverhalten, -Interessen, -Präferenzen
<b>Produktionsinformationen</b> (Anzahl produzierte Einheiten, Ausfallrate der Maschinen, Kosten, Energieverbrauch, Rückrufe)	Produktverbesserung, Verbesserung der Prozesse, Qualitätskontrolle



An allen Schnittstellen finden ETL-Prozesse statt

## 3.2 Logical model diagram



→ Relationale Datenbanken

- Zentral: Tables, IDs/Schlüssel, Spalten, Zeilen
- Verbindet **Tables** (vergleichbar Excel / SPSS sheets) über IDs
- Wideformat (eine Zeile = ein Fall) vs. Longformat (multiple Fälle genestet unter einer Variable)
- Abfrage mittels SQL (Structural Query Language), ist aber in R mittels des dbplyr package möglich
- Einfacheres (fiktives) Beispiel ([download](#) Excel-file)

- "Extract, transform, load"

- **Extract:** Zugriff auf tables und Variablen/features (`select`, `filter`)
- **Transform:**
  - **Integrieren** verschiedener Tables (z.B. `left_join`) anhand der ID oder anderer Variablen
  - **Data Cleaning**
    - **Outlier** entdecken und "behandeln" (z.B. Trimming oder Winsorizing) (`DescTools::Winsorize`)
    - **Duplicates** entdecken und eliminieren (`janitor::get_dupes`)
    - **Missing data** entdecken, verstehen und behandeln (z.B. imputieren) (siehe Tidyverse-Skript)
  - **Datenaufbereitung** (Data Wrangling / Feature Engineering) für die Analyse (→ vgl. Informationsstrategie)
    - **Rekodieren**, z.B. von strings in factors (`recode`, `mutate`)
    - **Mapping**, z.B. Verschiedene Versionen von Einträgen in ein unified value (M, male, Mr.)
    - **Berechnen**: Z.B: Sales = Anzahl verkaufter Teile \* Preis (`mutate`)
    - **Aggregieren**, z.B. Tagesumsätze auf Monatsumsätze, oder Regionalebene (`summarize`)
    - **Pivoting**, z.B. Wideformat in Longformat (`pivot_longer`)

- **Rohdaten:** Posts auf einem social media Kanal. Rohdaten wären...
  - Text: 1) Firmenrelevante Postings, 2) eigene Postings
  - Meta-Daten: Datum, Uhrzeit, Name, u.U. andere Infos (PLZ) etc.
- **Preprocessing:**
  - Tokenizing, stopwords removal, lemmatization
  - Ggfls. Match mit anderen Kundeninfos (DSGVO!)
- **Analyse:**
  - Z.B. Sentiment-Analyse mittels Match mit Wörterbüchern mit positiven/negativen Wörtern
  - Analytische Optionen
    - Individuelle Zeitreihe mit Sentiment-Verläufen (Trends? Systematische Rhythmen?)
    - Zusammenhänge mit anderen relevanten Variablen (Käufe, Kündigungen)?

- **Ausgangslage:**
  - Hypothesenbasiert oder
  - Explorativ / Data-Driven ?
- **Analytisches Ziele:** → Was ist die relevante / zu liefernde Information (Passung zum Informationsbedarf)?
  - **Deskriptiv**
  - **Kausal (Explanation)**
  - **Vorhersage (Prediction)**

- **Arten:**

- **Klassisches "Reporting"** (→ Business Intelligence)
- **Momentaufnahme der Leistung** (Beschreibung von KPIs mittels deskriptiven Maßen)
- **Trends** darstellen (z.B. Verläufe der Verkaufszahlen oder Zufriedenheit)
- **Muster entdecken** (z.B. homogene Kundengruppen entdecken)

- **Ansätze:**

- **Univariate / multivariate deskriptive Statistiken** (Mittel, Streuung, Korrelation, etc.)
- **Visualisierungen:** Verteilungen, Zusammenhänge (z.B. Scatterplot)
- **Zeitreihenanalyse:** Trends? Saisonale Schwankungen? (→ STL Decomposition)
- **Clusteranalyse** (→ "Unsupervised Machine Learning")



- **Merkmale:**

- Man hat Vermutungen über die Rolle bestimmter Variablen
- Ziel: Kausalen Effekt schätzen (als Lead oder Lag Information).
- Zentrales Kriterium: Korrektheit des Effekts

- **Beispiele**

- Warum kündigen Kunden?
- Welche Wirkung wird eine Veränderung des Preises haben?
- Welche Merkmale eines Produktes attrahieren Kunden?

- **Ansätze (klassisch!)**

- **Experimentell** (z.B. A/B Testing, Feldexperimente (randomisiert v.s. quasi-experimentell))
- **Längsschnitt- und Zeitreihenansätze:** VAR Models oder Interrupted Time Series
- **Regressionsanalyse** (inkl. logischer Regression, Survivalanalyse) mit Kontroll- oder Instrumentalvariablen

- **Merkmale:**

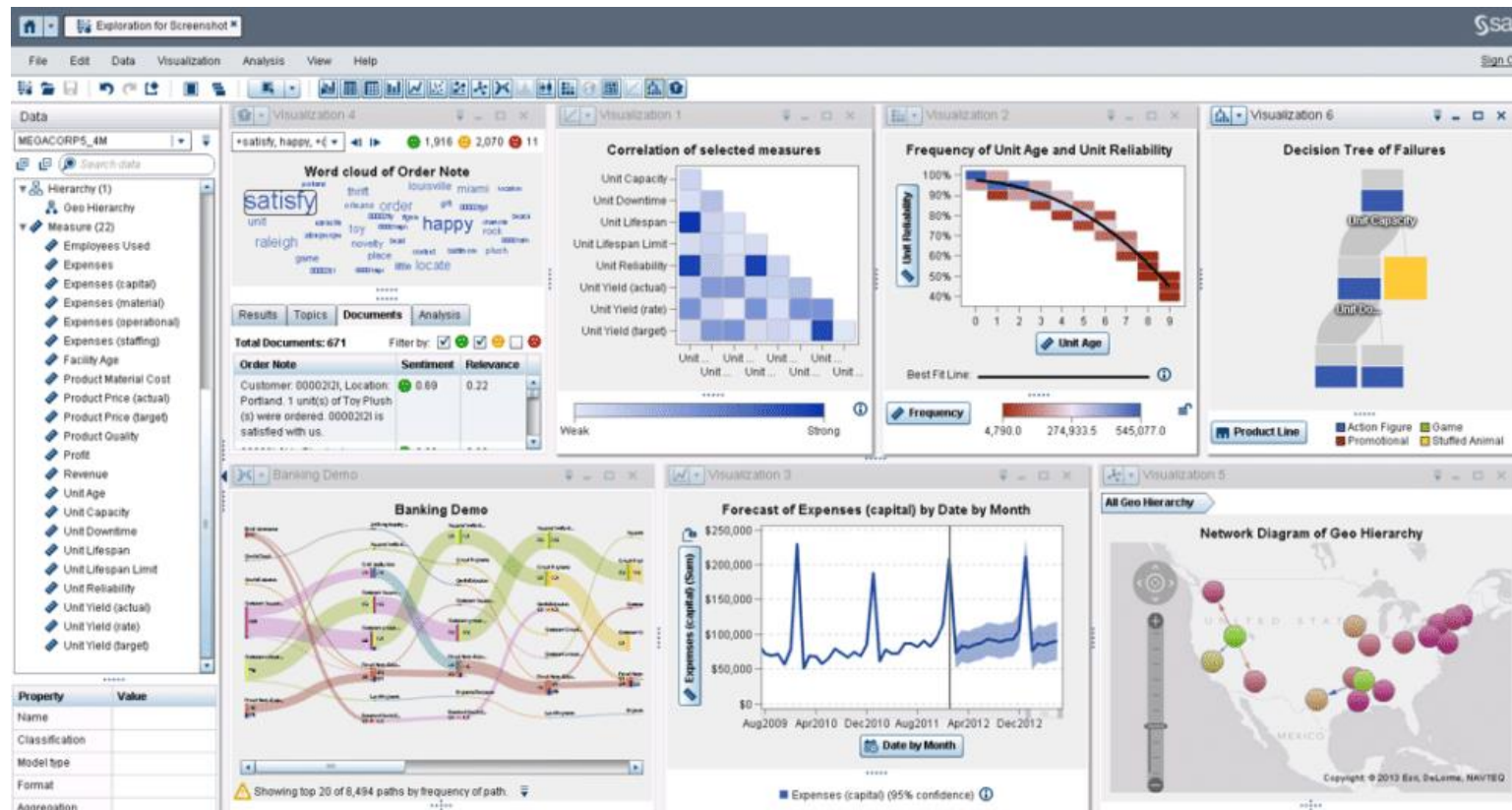
- Klassische Anwendung von Machine learning / KI
- Rein prädiktives (nicht-kausales Ziel): Was weiß ich über Y, wenn ich X kenne?
- Zentrales Kriterium: Geringer Vorhersagefehler

- **Beispiele:**

- Forecasting von Abonnement-Zahlen
- Vorhersage der Job Performance von Bewerbern
- Identifikation von Kennzeichen für Kreditkartenmissbrauch
- Market Segmentation mittels Clusteranalyse

- **Vorteil: Automatisierbarkeit** (z.B. Recommender Systems, Benachrichtigungen bei kritischen Beschwerden oder möglichen Maschinenausfällen)

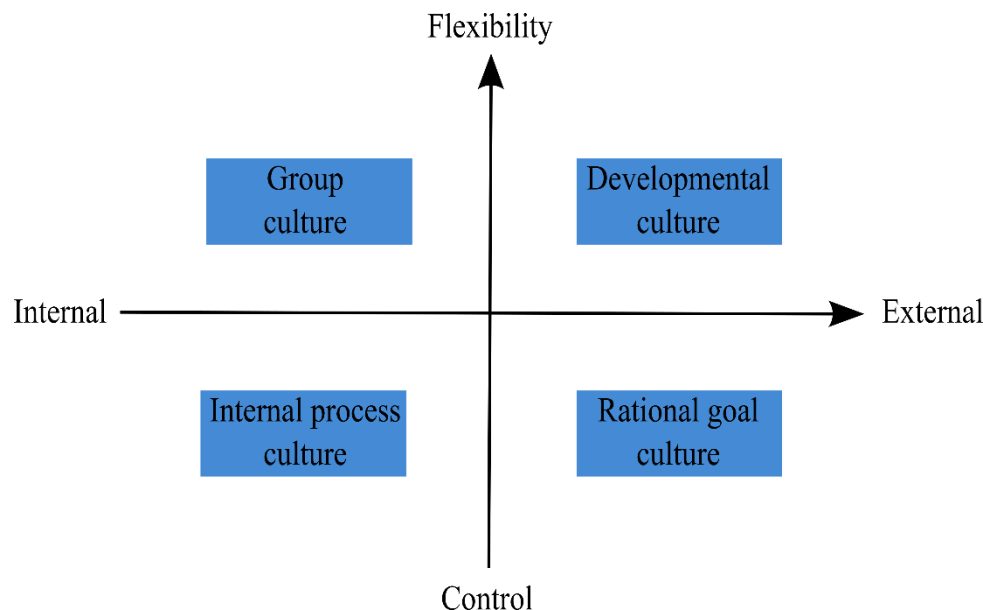
- Berichte? Präsentation?
- Model Deployment: *"Bringing the model into production"*:
  - Dashboards
  - Automatisierter KI-Prozess? Beispiel: Kunde bekommt eine mail mit einem Angebot, wenn er in "auffällig" wird





# Implementierung der BA- Funktion

- **Mittelstand:** Traditionell geprägte Strukturen und Kulturen
- **Orientierung: Kultur-Taxonomie** von Quinn & Rohrbaugh (1983)
  - Beschreibt das Verhältnis von Kultur, Organisationsstruktur und Umwelt
  - **Struktur:** Formalisiert & zentralisiert (*mechanistisch*) vs. wenig formalisiert / flexibel und dezentralisiert (*organisch*)
  - **Fokus:** Internal (auf interne Abläufe und Personen gerichtet) oder external (auf das Überleben der Organisation und die Umwelt gerichtet)



- **Group Culture:** Harmonie, Kohäsion, Mitarbeiterzufriedenheit ist wichtig
- **Internal Process Culture:** Klare Hierarchien, Formalisierung, Transparenz und Effizienz
- **Developmental Culture:** Dynamisch, innovativ, gute Fehlerkultur
- **Rational Goal Culture:** Ambitioniert, Wachstum, Langzeitorientiert, Effizienz

- Eine BA-Implementierung ist eine übliche **Organisationsentwicklungs-Maßnahme**
  - **Resistenzen bei Mitarbeitern** aufgrund
    - **Sorgen/Ängste** z.B. vor Überforderung, Arbeitsplatzverlust oder Privatheit (z.B. bei Einsatz im HR)
    - **Verständnisprobleme** (*"warum ist sowas nötig? Es ging doch bislang ohne?"*)
    - **Mangel an Beteiligung** (z.B. wenn Lösungen für User nicht an deren Bedarfe angepasst werden)
  - **Resistenzen bei Managern** aufgrund
    - **Angst vor Verlust an Entscheidungsmacht** (durch Delegation der Entscheidungsmacht an "Daten")
    - **Kognitive Dissonanz** (wenn Evidenz den eigenen Berufserfahrungen widerspricht)
    - **Mangel an Ressourcen**

- Aspekte bei der Implementierung
  - **Manager und Mitarbeiter** einbeziehen: Verstehen, wo sie Unterstützung brauchen, den Nutzen verdeutlichen
  - **Die zentralen Personen überzeugen** (das heißt Manager und einflussreiche Mitarbeiter) → Rollenmodelle vs. "Killer"
  - **Training und Unterstützung**
  - **Start small and make it happen:** Wichtigkeit anfänglicher kleiner erfolgreicher Projekte
  - **Implementierung begleiten und feedback einholen** (→ lag information über BA-Erfolg)

- Das BACC ist eine interdisziplinäre Gruppe, die die Implementierung und Steuerung von BA-Aufgaben koordiniert
- Besteht aus Vertretern der 3 Rollen (Business, Analytics, IT)
- **Aufgaben und Funktionen**
  - **Integration der Fachperspektiven**
  - **Identifikation und Beseitigung von Barrieren** (siehe vorheriger Aspekt)
  - **Identifikation von Informationsbedarfen und Planung von Projekten** (Datenverfügbarkeit, evtl. Erhebung, Design, Analytik), vgl. Abschnitt "Rolle der Analysten"
  - **Stärkung der Legitimität**
  - **Integration von Analysten verschiedener Funktionsbereiche**
  - **Identifikation von Entwicklungsbedarfen** was Technik und personelle Kompetenzen angeht (→ Einstellung oder Training bzgl. Analytik oder IT)



- Art der Implementierung

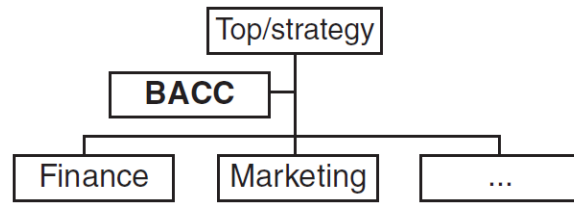
- als **formale Einheit**:

- Separate Abteilung mit fest angestellten Mitarbeitern
    - V.a. in großen Firmen

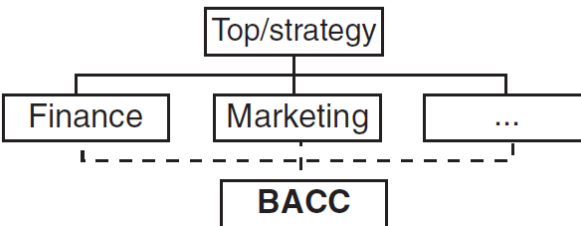
- als **virtuelle steering group**:

- Vertreter der Funktionsbereiche
    - Nachteil: Nebenjob
    - V.a. in SMEs (Es fehlen Ressourcen für eine formale Einheit)

Formal Organizational Unit



Virtual Organizational Unit



- **Veränderungen der Umweltbedingungen** → Komplexität, Dynamik, Unsicherheit
- **Digitalisierung und Big Data** erhöhen Datenverfügbarkeit und Herausforderungen
- **"BA-Funktion":**
  - System von Aktivitäten, mittels Datenauswertung Entscheidungen mit operativer oder strategischer Relevanz zu unterstützen
  - Mehr oder weniger stark an strategischen Zielen (auf Unternehmens- oder Funktionsbereichsebene) orientiert (von isoliert bis integriert)
- **Integration von 3 relevanten Rollen:** Business, Analyst, IT
- **Kategorien von Informationen und Analysezielen:**
  - Produktbezogen
  - Kundenbezogen
  - Prozessbezogen

- **Besondere Rolle des Analysten** als Brückenbauer und Implikationen für Aufgaben und Kompetenzen
- **Konzept der Informationsstrategie:**
  - Klärung des Informationsbedarfs / Informationsziels
  - Klärung der zentralen relevanten Phänomene
  - Identifikation / Generierung relevanter Daten
  - Klärung des passenden analytischen Designs und des Aufwands für Cleaning / Feature Engineering
  - Klärung der intendierten Nutzung der Ergebnisse
- Rolle des Data Warehouses als Infrastruktur zum Sammeln, Integrieren, Aufbereiten und Verfügbar-machen von Daten
- Arten und mögliche (kulturelle/strukturelle) Probleme bei der Implementierung einer BA-Funktion