

SPARBUCH

Vor langer Zeit gab es exponentielles Wachstum ... auf dem Sparbuch



Bildquelle: Konstantin Evdokimov/Unsplash

The sexy job in the
next ten years will
be statisticians.



DATEN- & KI-KOMPETENZEN
Future Skills for Future
Leaders

Katharina Schüller, CEO STAT-UP
katharina.schueller@stat-up.com



AGENDA

[[1]] Smart Data & Analytics =
Erfolgsfaktoren

[[2]] Data & AI Literacy =
Kompetenzen

[[3]] Nachhaltige & integrierte
Datenstrategie



[] Smart Data & Analytics =
Erfolgsfaktoren

- Daten wachsen exponentiell, Verdopplungszeit 2-3 Jahre
- Begriff „exponentielles Wachstum“ weithin bekannt und kaum verstanden
- Wichtige Erkenntnis: „mehr Daten“ ungleich „mehr Wissen“
- Pandemie: Mangel an guten Daten
- Hauptrolle von Daten- und KI-Experten dabei: schlechte Daten „reparieren“, z.B. Lücken schließen

→ DATEN SIND GRUNDLAGE FÜR ENTSCHEIDUNGEN, DARIN LIEGT IHR WERT



Geschäftsmodelle von Banken stehen auf dem Prüfstand → datengetrieben?

THESE: Daten = Öl des 21. Jahrhunderts

1. Daten sprudeln wie Öl
2. Daten wie Öl gibt es in zahlreichen Formen und Anwendungsmöglichkeiten
3. ABER: Daten sind bloß Rohmaterial

Nicht Rohmaterial zählt, sondern Wertpotenzial

Wertpotenzial ungleich Produkt (Benzin), sondern Bedürfniserfüllung (Mobilität)

→ DATEN SIND NICHT DAS ZIEL, GENAUSO WENIG WIE STATISTIKEN





→ ZIEL = WERTBEITRAG VON DATEN FÜR
BESSERE ENTSCHEIDUNGEN

- Wertschöpfung aus Daten misslingt häufig
 - Ca. ¾ aller Analytics-Projekte scheitern
 - 90% der Daten in Unternehmen liegen brach
 - Kosten schlechter Datenqualität mehrere 100 Mrd USD, allein in USA
- **SMART DATA** = qualitativ hochwertige Daten, direkt verarbeitbar zu entscheidungsrelevanter Information
- **ANALYTICS/AI** = Werkzeuge, um Maximum an Information aus Daten zu gewinnen



[[2]] Data & AI Literacy =
Kompetenzen

Datenkompetenz = Voraussetzung für
Relevanz/Nutzen Analytics-Ergebnisse

1. Für nachhaltige Ausrichtung an
Kundenbedürfnissen

Kunden besser verstehen, passgenaue,
datenbasierte Produkte/Services entwickeln

2. Für strategische Entscheidungen

Neue Märkte/Kundengruppen festlegen

3. Für Verbesserung Wirtschaftlichkeit

Angebot optimieren, präzise Risikobewertung,
operative Leistung steigern

→ DATENKOMPETENZ IST ZENTRALER
BAUSTEIN EINER DATENSTRATEGIE

Datenkompetenz („DATA LITERACY“) wird
Schlüsselkompetenz für alle

→ 4 Fragen in Zusammenhang mit der
Geschäftstätigkeit einer Bank:

1. Was **WILL** ich mit Daten machen?
 2. Was **KANN** ich mit Daten machen?
 3. Was **DARF** ich mit Daten machen?
 4. Was **SOLL** ich mit Daten machen?
- Datenethik wird vom bloßen Schlagwort zu konkreter Roadmap
 - Auch KI-Kompetenzen werden eine immer größere Rolle spielen



Unerlässlich: **VERSTÄNDNIS** von

- Einfluss der Technologien auf Interaktion von Menschen
- Chancen & Grenzen von Daten/Algorithmen

Größtes Problem:

- Blindes **VERTRAUEN** von Menschen in Black-Box-Algorithmen
- Mangelnde menschliche Kompetenz (in doppelter Hinsicht), Entscheidungen von Maschinen zu hinterfragen

“It’s the wild west right now.
People try to twist and turn
and use every tool to get
sense out of these data sets.
But we’re flying a little bit
blind at the moment.”

GOOGLE FLU TRENDS

- Grippewellen früher vorhersagen aus Suchanfragen nach Grippesymptomen
- Umkehr Ursache - Wirkung: Grippewelle um 50% überschätzt

CHURN PREVENTION

- Korr. Vertragslaufzeit - Kündigung
- Kündiger: kürzere Laufzeiten, weil sie gekündigt hatten
- Nicht-Kündiger waren zum Zeitpunkt der Analyse noch Kunden

→ FALSCHER SICHT AUF DIE DATEN FÜHRT ZU TRUGSCHLÜSSEN

INTERNET-NUTZUNG

- Online-Befragung (2016): Können Sie sich Leben ohne Internet vorstellen?
- Indien 82%, Japan nur 62%
- Aber: Internetnutzung Japan 91%, Indien 28%
- D.h. 77% der Inder können/müssen sich Leben ohne Internet vorstellen

SCORING AUS BESTANDSKUNDEN

- Zusammenhänge mit bestimmten Merkmalen > Rekonstruktion früherer Scorings?
- MAN LERNT, WAS MAN SCHON WEISS, bzw. dass Kampagne funktioniert hat

TARGET

- Klage: Werbung verführe 17-jährige Tochter zum Schwangerwerden
- Aber Tochter war schon schwanger
- Target: Algorithmen identifizieren schwangere Kundinnen wg. veränderten Kaufverhaltens
- Aufschrei > Privatsphäre?!?

→ LAW OF TRULY LARGE NUMBERS

- Daten von über 100 Mio Amerikanern
- Selektion aller Frauen im gebärfähigen Alter > zu jedem beliebigen Zeitpunkt 100.000+ jetzt bzw. in 1-2 Wochen schwanger

FACEBOOK/TRUMP-KAMPAGNE

Orientierung zu 88% aus Likes
vorhersagbar - was heißt das?

Annahme: 10% homosexuell, d.h. von 10.000
Personen > 9.000 hetero-/1.000 homosexuell

880 richtig klass. Homosexuelle

1.080 falsch klass. Heterosexuelle

- Treffer $880 / (880 + 1.080) = 45\%$

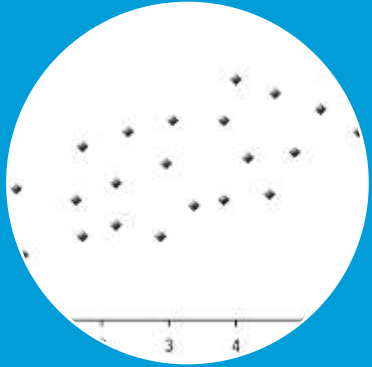
PROBLEM: VIELE KOMBIS VIELER DATEN

- Microtargeting kurzfristig billig
- Bis zu 100.000 A/B-Tests/Tag
Signifikanzniveau 5%: 5.000 f.p./Tag

→ LANGFRISTIGER SCHADEN
(Wirkungsverlust von Kampagnen)



“There are a lot of small data problems that occur in big data. They don’t disappear because you’ve got lots of the stuff. They get worse.”



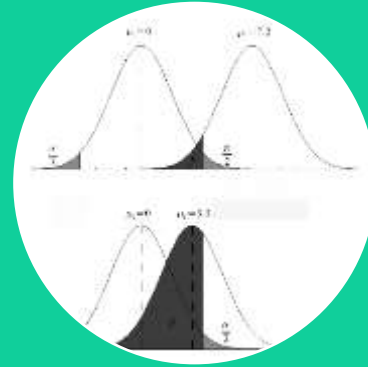
CORRELATION VS. CAUSATION

Ohne Idee, wie es zu
der Korrelation kommt,
kann man nicht wissen,
was diese Korrelation
zerstören könnte



SAMPLING BIAS

Wer jede Nachricht auf
Twitter analysiert,
erfasst trotzdem nicht
die Stimmung der
Weltbevölkerung.
Ethische Komponente:
Algorithmic Bias



FALSE POSITIVES

Insbesondere bei
seltenen Merkmalen ist
das Risiko von False
Positives enorm hoch,
selbst wenn Algorithmen
hohe Sensitivität und
Spezifität besitzen



ERROR INFLATION

Tausende Merkmale
werden auf
Zusammenhänge
untersucht -
irgendwo wird immer ein
Effekt gefunden





DATENBEZOGENE HERAUSFORDERUNGEN

Datenbeschaffung, Datenerfassung,
Datencharakteristika, Datenqualität,
Datenbereinigung



PROZESSBEZOGENE HERAUSFORDERUNGEN

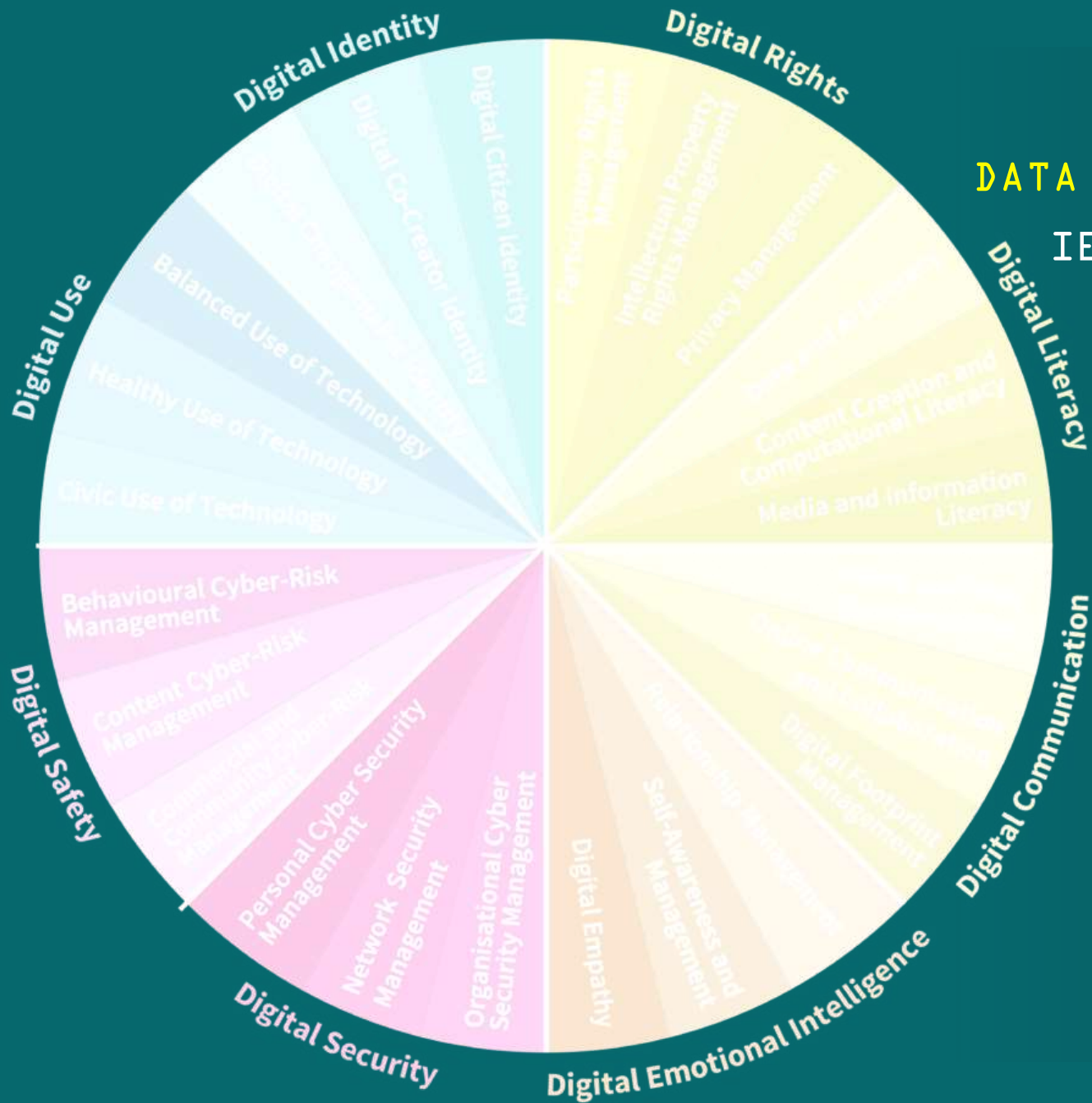
Speicherung, Integration, Transformation,
Modellwahl, Implementierung,
Interpretation, Ergebnispräsentation



MANAGEMENT-HERAUSFORDERUNGEN

Privatsphäre, Sicherheit, Steuerung,
Ethische Aspekte

“Big data has arrived,
but big insights have not.”



DATA & AI LITERACY

IEEE 3527.1: DIGITAL INTELLIGENCE DQ

Ability to

- generate, process, analyze, present **MEANINGFUL INFORMATION** from data
- and develop, use, and apply AI and related algorithmic tools and **STRATEGIES**
- in order to guide informed, optimized, and contextually relevant **DECISION-MAKING** processes

KODIEREN

DEKODIEREN



Reale Welt



BEOBACHTEN



HANDELN



Daten

Wissen



MODELLIEREN

INTERPRETIEREN



Information

Entfernen von Kontext

Kontext hinzufügen



DER WERTSCHÖPFUNGSPROZESS

[1] Rohmaterial → DATEN

Edelstein + Edelmetall

[2] Aufbereiten → INFORMATIONEN

Schmieden/Schleifen

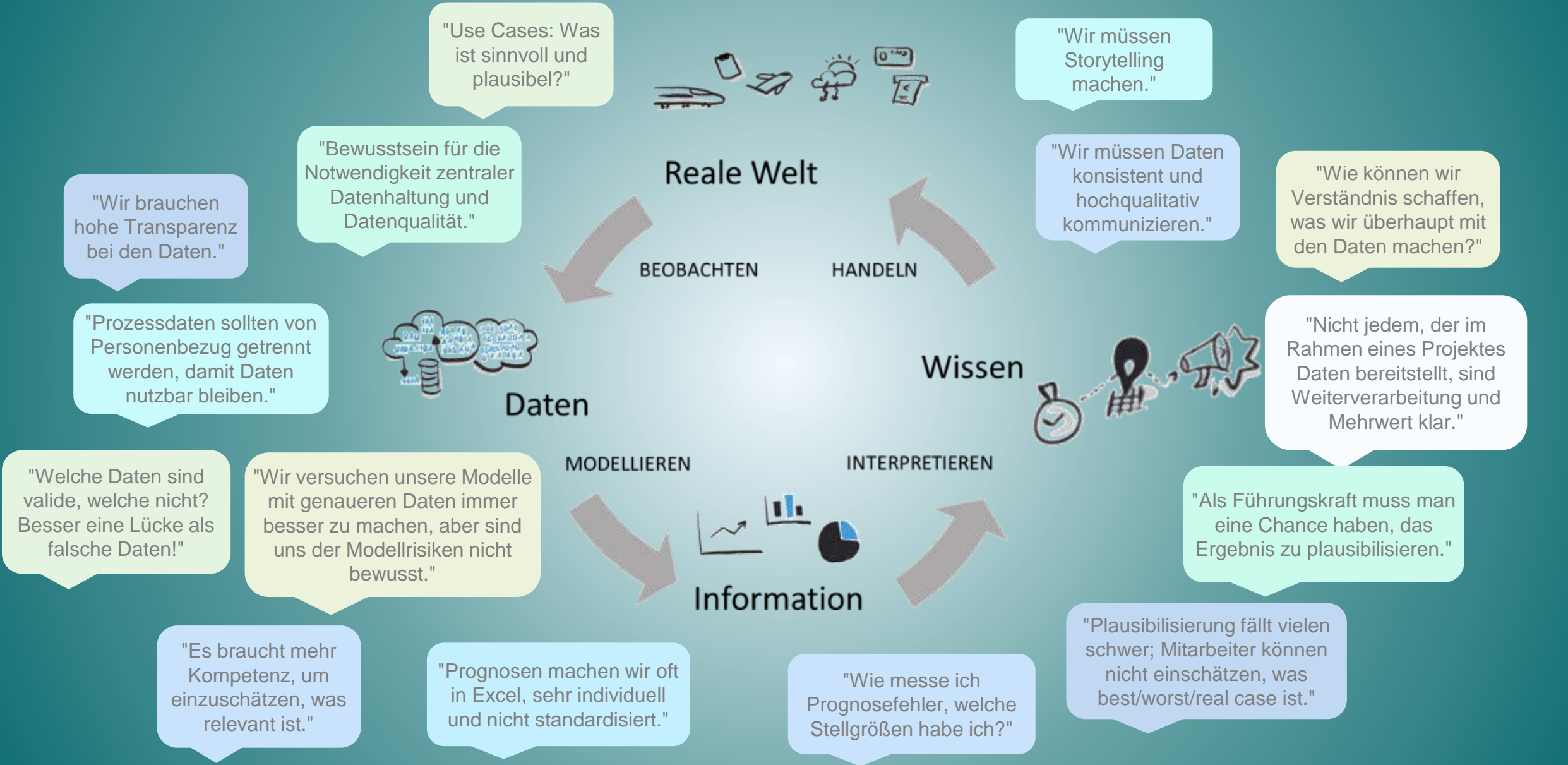
[3] Kombinieren → WISSEN

Ring

[4] Bedeutung verleihen/
interpretieren → HANDELN/MACHT

Tiffany's Stempel

→ Größter Teil der Wertschöpfung
passiert in Schritt 4 und steckt nicht
im Rohmaterial!



"Use Cases: Was ist sinnvoll und plausibel?"

"Wir müssen Storytelling machen."

"Wir brauchen hohe Transparenz bei den Daten."

"Bewusstsein für die Notwendigkeit zentraler Datenhaltung und Datenqualität."

"Wir müssen Daten konsistent und hochqualitativ kommunizieren."

"Wie können wir Verständnis schaffen, was wir überhaupt mit den Daten machen?"

"Prozessdaten sollten von Personenbezug getrennt werden, damit Daten nutzbar bleiben."

"Nicht jedem, der im Rahmen eines Projektes Daten bereitstellt, sind Weiterverarbeitung und Mehrwert klar."

"Welche Daten sind valide, welche nicht? Besser eine Lücke als falsche Daten!"

"Wir versuchen unsere Modelle mit genaueren Daten immer besser zu machen, aber sind uns der Modellrisiken nicht bewusst."

"Als Führungskraft muss man eine Chance haben, das Ergebnis zu plausibilisieren."

"Es braucht mehr Kompetenz, um einzuschätzen, was relevant ist."

"Prognosen machen wir oft in Excel, sehr individuell und nicht standardisiert."

"Wie messe ich Prognosefehler, welche Stellgrößen habe ich?"

"Plausibilisierung fällt vielen schwer; Mitarbeiter können nicht einschätzen, was best/worst/real case ist."



[[3]] Nachhaltige & integrierte Datenstrategie

„Gläserner Kunde“ = Illusion

- In den Kunden „hineinschauen“ heißt nicht, das Richtige zu sehen und die richtigen Schlüsse zu ziehen

→ **ES GEHT NICHT UM DATEN** (Kunden-/Produkt-/Vertriebsdaten), **SONDERN UM MUSTER IN DEN DATEN**

- Gezielter Einsatz von Data Analytics & selbstlernender Algorithmen (Machine Learning)

Z.B. Anwendungen entwickeln, die automatisiertes, datengetriebenes Marketing für eine bestimmte Kundenzielgruppe oder Produktgruppe per E-Mail oder Brief steuern



Ziele der integrierten Datenstrategie:

1. Datenkultur zur Entwicklung und Implementierung von Anwendungsfällen schaffen
2. Kompetenzen aller Akteure dauerhaft auf konkurrenzfähige Produkte und Dienstleistungen ausrichten

Fähigkeit der Organisation, Daten systematisch in Wert zu verwandeln.

Betrifft Analyseprozess an sich (Coding) und Interpretation von Ergebnissen (De-Coding)

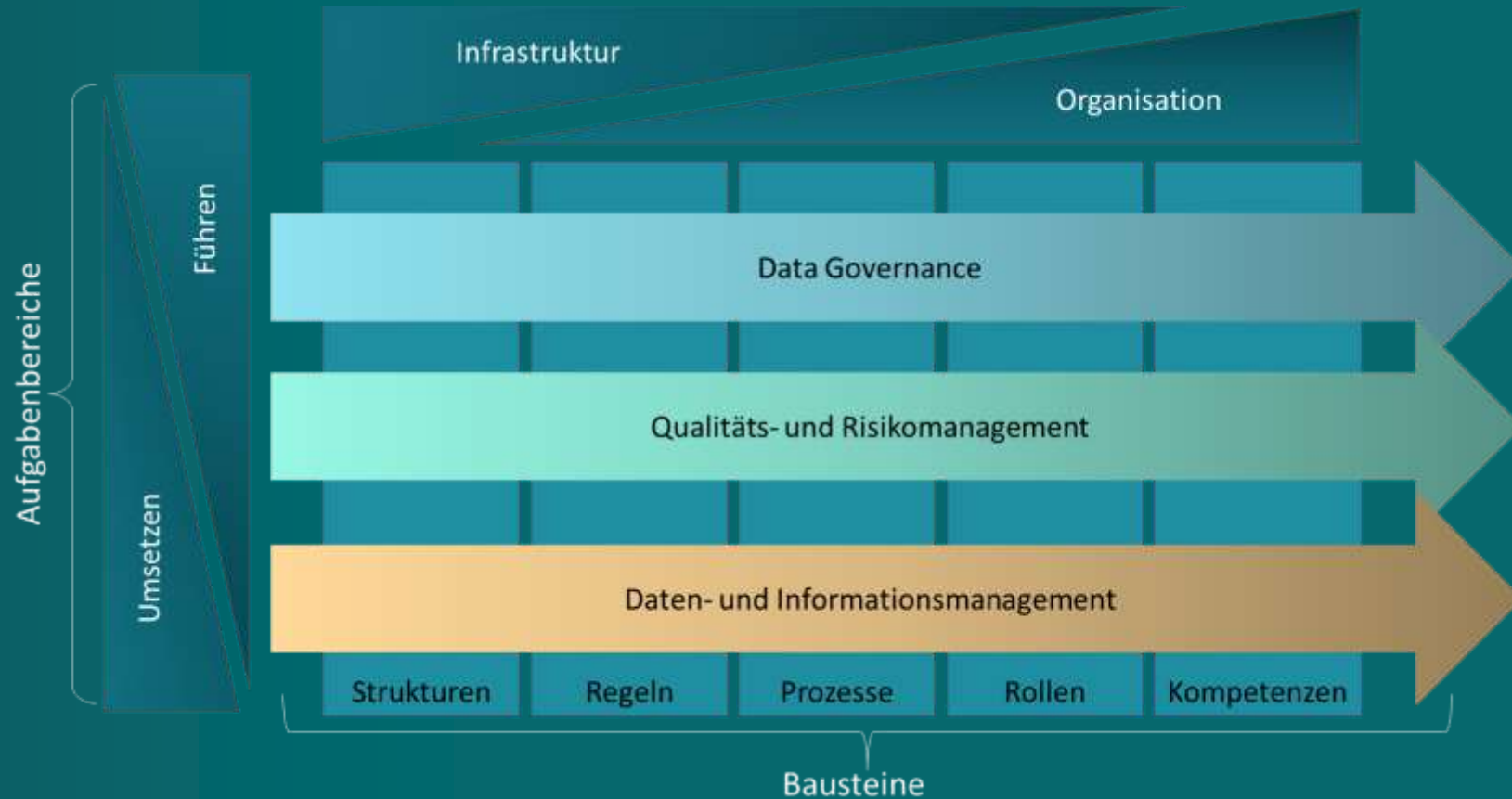
→ Interpretation ist nicht Teil der Daten - **DATEN UND BEDEUTUNG VON DATEN SIND WIE RING UND STEMPEL**

„Daten-DNA“ durchdringt

- Strukturen
- Regeln
- Prozesse
- Rollen
- Kompetenzen

Kultur-Change betrifft

- Organisation
- Infrastruktur
- Führung
- Umsetzung



→ Rahmenbedingungen schaffen, um
DATEN AUS SILOS HERAUSZUHOLEN



FAZIT

Datenstrategie muss Organisation wie ein Kompass neu ausrichten und drei Fragen konsistent beantworten:

1. Was wollen und können wir zukünftig mit Daten machen?
2. Wie machen wir es richtig?
3. Und wie machen wir das Richtige?

→ Ethische Parallele von Daten und Öl:

Inkompetenter Gebrauch/Missbrauch führt zu erheblichen Schäden

Öl: Verkehrskollaps, Feinstaub, Plastikmüll in den Meeren

Daten: Verlust von Selbstbestimmung, Diskriminierung, falsche Entscheidungen

SPARBUCH

GELD verdient man heute mit **DATEN**.

Darauf müssen sich Banken vorbereiten - heute, nicht morgen.

CHANCE: Den Menschen nicht aus dem Blick zu verlieren

→ Wie wollen wir zukünftig **WERT** schaffen, ohne **WERTE** zu verletzen?

Ethik & (Wert-)Haltung garantieren kein exponentielles Wachstum, aber ein nachhaltiges - das, was wir uns einmal vom Sparbuch erhofft haben

SPARBUCH

We have a new resource here. But nobody wants „data“. What they want are the answers.