

WIRTSCHAFTSINFORMATIK

VO3 — DATENARTEN, DATENMODELLE & DATENQUALITÄT

PROF. DR. CHRISTIAN BOCKERMANN

PROF. DR. JULIAN GROTE

PROF. DR. VOLKER KLINGSPOR

HOCHSCHULE BOCHUM · FACHBEREICH WIRTSCHAFT

WINTERSEMESTER 2026/27

- Frage: *Was passiert zwischen Klick auf „Bestellen“ und der Zahl im Monatsbericht?*
- Antwort: Die Bestellung durchläuft **viele Systeme** und wird in jedem **umgeformt**. Der Bericht zeigt das *Ergebnis*, nicht das Original.
- Genau deshalb heute die Anschlussfrage: Wenn jedes System umformt — können dann zwei Systeme **dieselbe** Sache *unterschiedlich* zählen?



Aktivierung — Kahoot: Kurzes Aufwärm-Quiz zur letzten Woche — ein paar Fragen, dann steigen wir ein.

DER VERTRIEB MELDET 1.616 BESTELLUNGEN, DAS CONTROLLING 1.594. WER HAT RECHT?



Aktivierung — Schätzung / Handzeichen: Woran liegt die Differenz von 22 Bestellungen — Rechenfehler, unterschiedliche Definition, oder kaputte Daten? Tippen Sie, bevor wir nachsehen.

Strukturiert — passt in Zeilen und Spalten:

- die Bestelltabelle, das ERP, jede CSV
- maschinell auswertbar *ohne* Vorarbeit

Unstrukturiert — kein festes Raster:

- E-Mails, Retouren-Fotos, Chatverläufe
- steckt voller Information, aber schwer zu zählen

Metadaten

Daten über Daten. Wann erfasst? Von welchem System? In welcher Einheit? — Ohne sie weiß niemand, was eine Zahl *bedeutet*.

Die Bestelltabelle ist strukturiert — und trotzdem streiten sich zwei Abteilungen über ihre Zeilen.

Quelle: Gabler Wirtschaftslexikon, Stichworte „strukturierte Daten“ und „Metadaten“, <https://wirtschaftslexikon.gabler.de> (abgerufen 07/2026)

- **Vollständigkeit** — fehlt etwas?
leere Kundennummer, kein Bestelldatum
- **Konsistenz** — widerspricht sich etwas?
dieselbe Bestellung in zwei Systemen
verschieden
- **Genauigkeit** — stimmt es mit der Realität?
Testbestellung über 0€, Storno als negativer Wert
- **Aktualität** — ist es noch gültig?
umgezogene Kundin, alte Adresse

Der Satz, den Sie mitnehmen

Eine Zahl ohne ihre **Definition** ist wertlos. „1.616 Bestellungen“ heißt gar nichts, solange nicht feststeht, *was als Bestellung zählt*.

Quelle: Gabler Wirtschaftslexikon, Stichwort „Datenqualität“, <https://wirtschaftslexikon.gabler.de> (abgerufen 07/2026)

Fast nie durch Schlamperei — meist durch den *Prozess*.

- **Doppelerfassung** beim Systemwechsel — dieselbe Bestellung wandert zweimal ins Berichtssystem.
- **Testbestellungen** im Produktivsystem — jemand prüft den Shop, die Zeile bleibt stehen.
- **Stornos** — eine Rücknahme steht als negativer Wert in derselben Tabelle.
- **Verwaiste Verweise** — eine Kundennummer, zu der es keinen Kunden (mehr) gibt.
- **Unterschiedliche Stichtage / Definitionen** — Vertrieb zählt alle Zeilen, Controlling nur bezahlte.

Deshalb ist „Wer hat recht?“ die falsche Frage. Richtig: „Was genau zählt jede Seite?“

1. Beide Zählungen reproduzieren.

Erst die 1.616 (alle Zeilen), dann sehen, wie man auf 1.594 kommt.

2. Die Differenz Zeile für Zeile lokalisieren.

Doppelte, Testbestellungen, Stornos, verwaiste Nummern — welcher Fehler wie oft?

3. Entscheiden, welche Definition gilt — und sie aufschreiben.

„Wir zählen als Bestellung: ...“ — damit die Zahl morgen wieder herauskommt.

Genau diesen Dreischritt machen Sie gleich in der Lernwerkstatt selbst.

Antwortsatz

Beide rechnen richtig und meinen Verschiedenes — der Unterschied sind **Testbestellungen und Doppelerfassungen**. Eine Zahl ohne ihre *Definition* ist wertlos.

- Der Vertrieb zählt *jede Zeile* (1.616), das Controlling nur *gültige, eindeutige* Bestellungen (1.594).
- Keiner hat sich verrechnet — sie haben nur nie festgelegt, **was zählt**.

Sie zählen zuerst selbst — mit einer *Schleife*, so wie man grundsätzlich zählt:

```
anzahl = 0
for wert in bestellwerte:      # ~20 Werte von Hand
    anzahl = anzahl + 1
print(anzahl)
```

... und merken: für 1.616 Zeilen will das niemand tippen. Dann übernimmt **pandas**:

```
len(df)                # 1616  -> alle Zeilen (Vertrieb)
df["bestellnr"].nunique() # 1608  -> eindeutige Bestellungen
```

- Daten sind **strukturiert** oder **unstrukturiert**; **Metadaten** sagen, was eine Zahl bedeutet.
- Datenqualität hat vier Fragen: **Vollständigkeit, Konsistenz, Genauigkeit, Aktualität.**
- Fehler entstehen meist im **Prozess**, nicht durch Schlamperei.
- **Eine Zahl ohne ihre Definition ist wertlos.**

Meilenstein

Mit dieser Woche sind **M1** (erste Programmschritte) und **M2** (Daten prüfen) abgeschlossen — der Selbsttest dazu liegt bereit.

Nächste Woche: Eine Kundin heiratet und zieht um — was ändert sich wo? Wir verbinden Tabellen.

- **Gabler Wirtschaftslexikon** — Stichworte „Datenqualität“, „Metadaten“, „strukturierte Daten“, <https://wirtschaftslexikon.gabler.de> (abgerufen 07/2026).
- **DAMA International** — *DAMA-DMBOK: Data Management Body of Knowledge* (2. Aufl., 2017), Kap. Data Quality — Dimensionen Vollständigkeit, Konsistenz, Genauigkeit, Aktualität.
- **Brick-Flow AG** — Fallunternehmen des Kurses, <https://brick-flow.ag/>. Der Rohdatensatz `brickflow_bestellungen_roh.csv` und die Zahlen 1.616 / 1.594 sind didaktische Konstruktionen (dokumentiert in `use-case/BRICKFLOW.md` §9.9).

Hinweis für Lehrende: Web-Quellen und Abrufdaten vor dem Einsatz kurz gegenprüfen — Prüfbarkeit ist Kursprinzip.