

m[method] 2 data

Data Warehouse
Einfach mit Methode

16. Tagung der DDVUG am 27.3.2025 in Wien

Data Warehouse – Einfach mit Methode

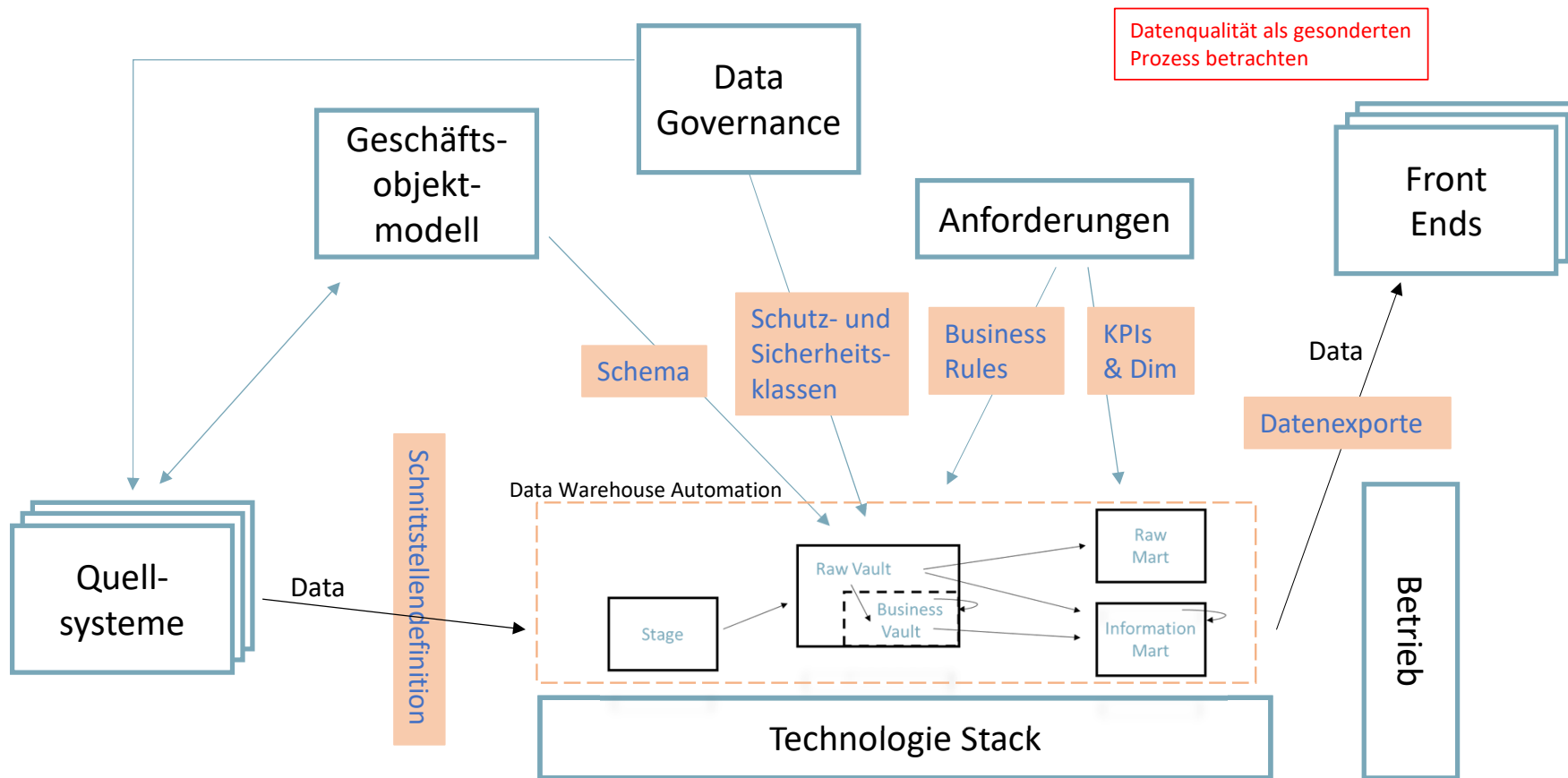
Warum?

- ★ Grundfrage
„Wie sieht ein gutes Data Warehouse aus?“
- ★ Andrew Hunt, Der Pragmatische Programmierer:
„Der Hund hat meinen Source Code gefressen“
- ★ All die Automation
die Arbeit wird weniger
die Dauer bleibt gleich
- ★ Das Data Warehouse wurde
noch nicht ganz so oft für tot erklärt, wie SQL

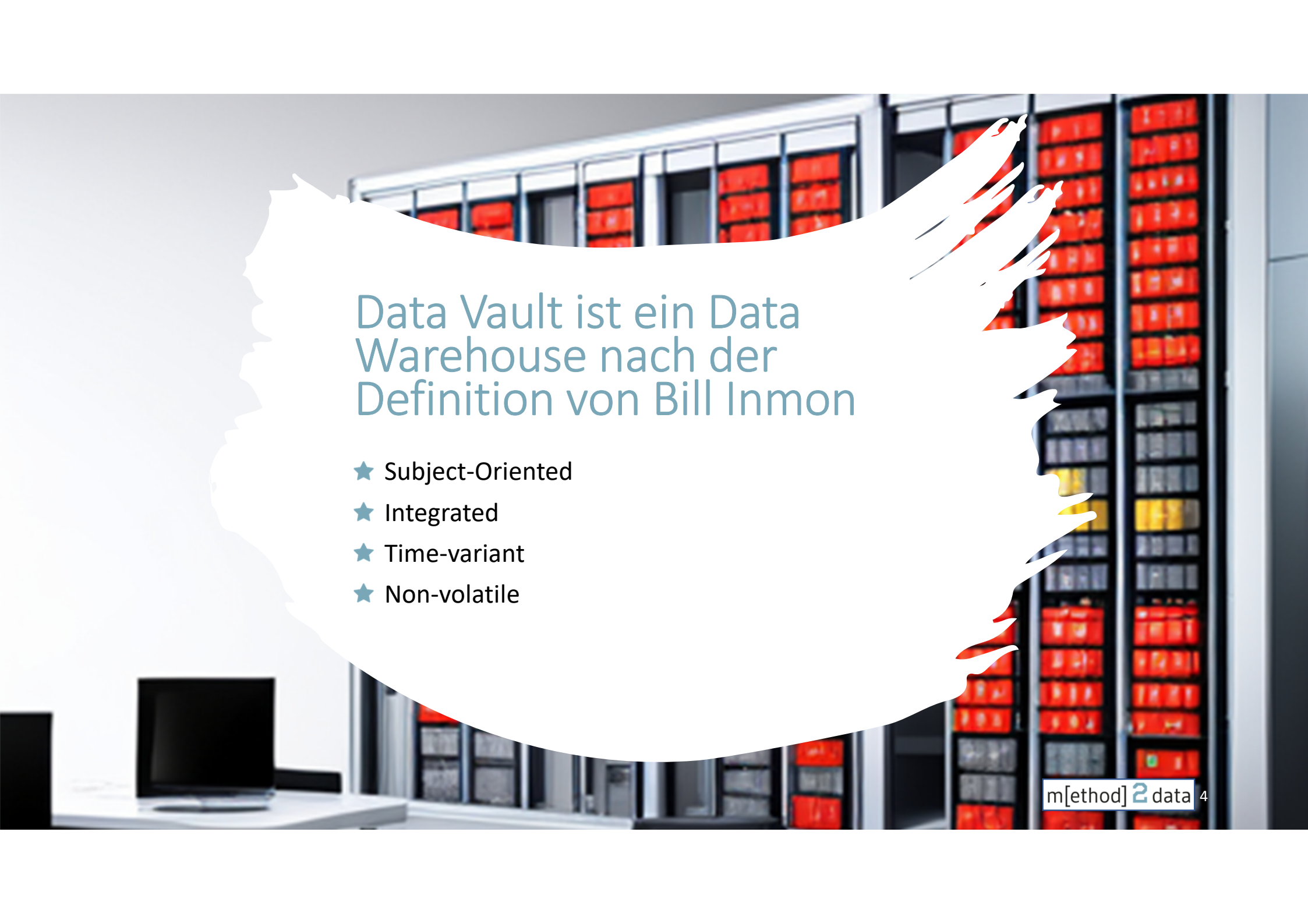
Ziele!

- ★ Skalierbar
von 1 Quellsystem bis hunderte
- ★ Schnell
- ★ Wächst mit den Aufgaben
 - ★ Anpassungsfähig
bei Änderungen in den Datenstrukturen
 - ★ Nachhaltig
dauerhafter Speicher der Unternehmens
transaktionen
- ★ Minium an ungeplanter Arbeit

Warum ist ein DWH komplex? Was frisst die Zeit?



- ★ 7 Schnittstellen sind viele und zeitraubend.
- ★ 5 der 7 Schnittstellen haben eine gemeinsame Basis.

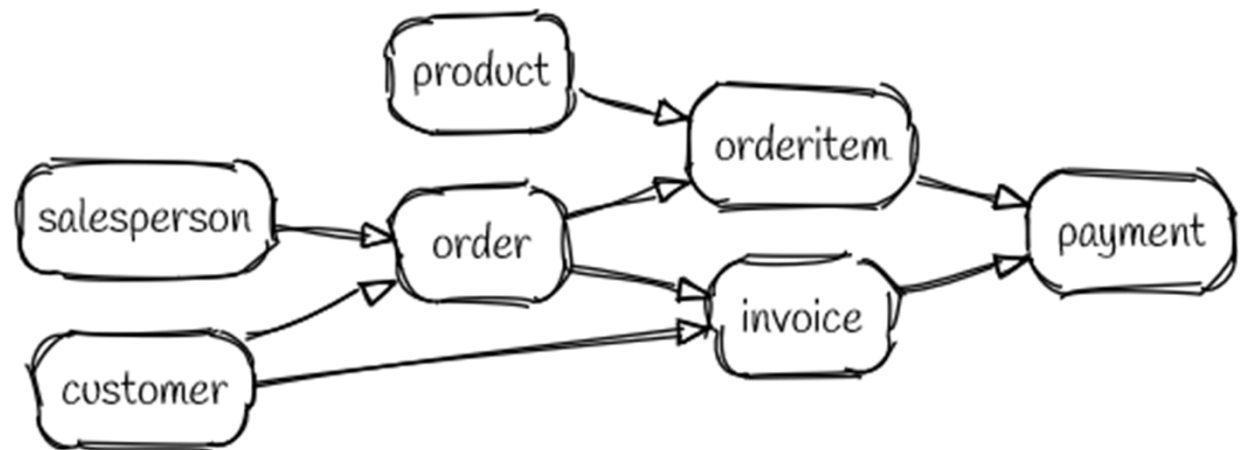


Data Vault ist ein Data Warehouse nach der Definition von Bill Inmon

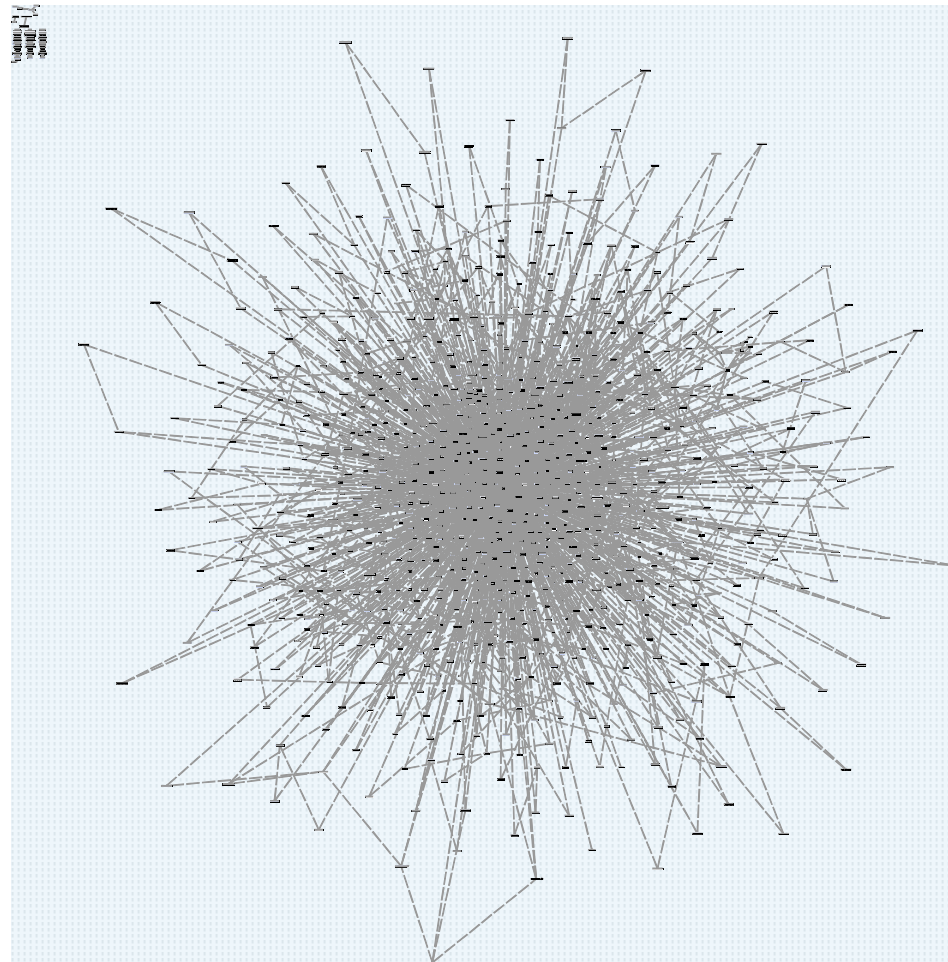
- ★ Subject-Oriented
- ★ Integrated
- ★ Time-variant
- ★ Non-volatile

Um Daten
verwalten zu
können,
müssen alle
Beteiligten
wissen,
welche Daten
verwaltet
werden

- Ein Unternehmensdatenmodell ist viel zu groß, um es in einem Zug zu verstehen
- Abstraktion ist der Schlüssel
 - ein konzeptionelles Modell verwenden
 - bestehend aus zentralen Geschäftskonzepten und deren Beziehungen
 - nur einen relevanten Teil des Modells betrachten
 - Geschäftsbegriffe verwenden
 - das Modell in einer lesbaren Form zu formatieren



Warum haben Datenmodelle so einen schlechten Ruf?



Ziel und Grenzen bei der Geschäftsobjektmodellierung

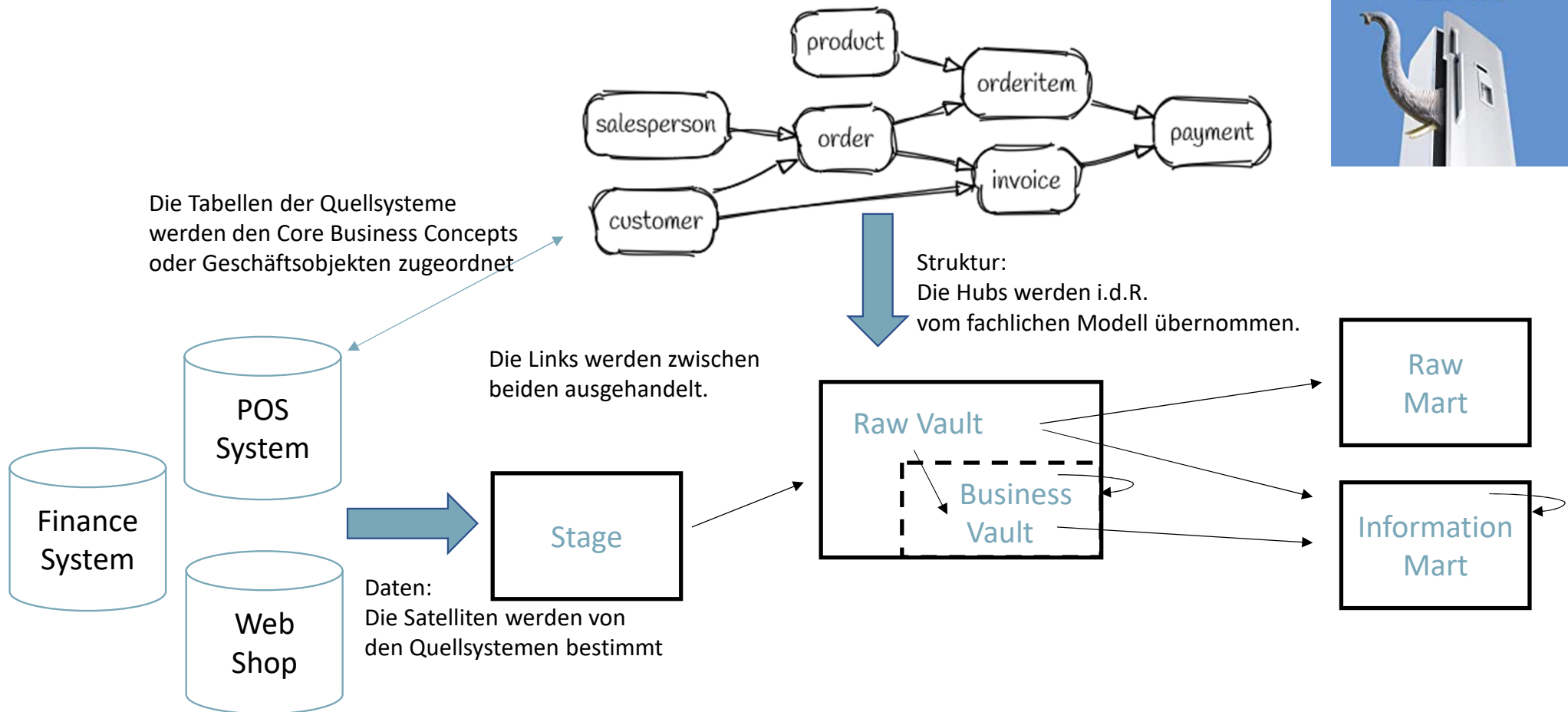
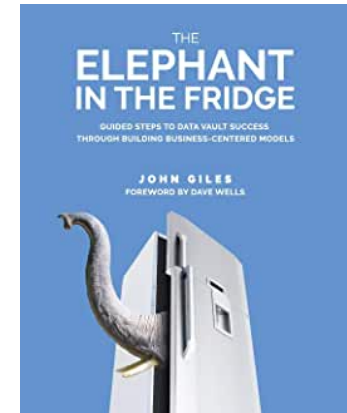
Ziel

- ★ mit Geschäftsobjektmodell und Glossar ein gemeinsames Datenverständnis schaffen
- ★ das Glossar als Basis für Requirements
- ★ individuelle, abweichende und notwendige Sichtweisen bzw. Blickwinkel abbilden
- ★ Klärung der Frage:
'Welche Daten haben wir?'
- ★ Es muss lediglich 'Gut genug' sein

Out of Scope

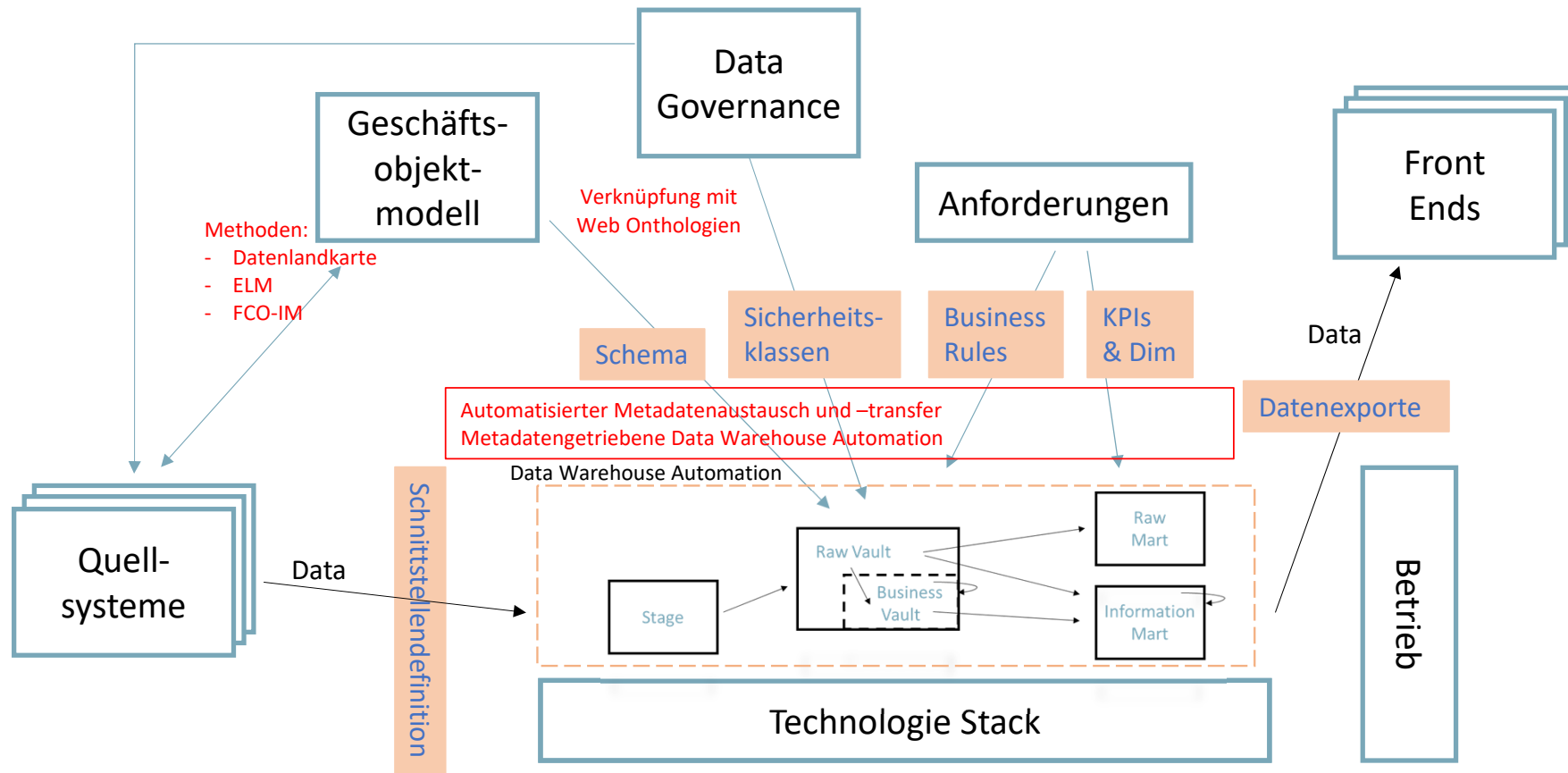
- ★ Enterprise Data Model
- ★ Universe of discourse
- ★ endlose Abstimmungen über den richtigen Begriff
- ★ alles was nicht direkt zu einem Data Warehouse führt

Integration bei Data Vault



Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

Datenqualität als gesonderten Prozess betrachten



Schnittstellenvereinbarung



Klare Vereinbarung,
service level agreement (SLA)



Umfang der Daten
klar beschrieben



Rechtzeitige Kommunikation
der Änderungen am Operativen
als Teil des SLA



Nicht-funktionales ebenfalls
vereinbaren: Performance,
Sicherheit und Skalierbarkeit

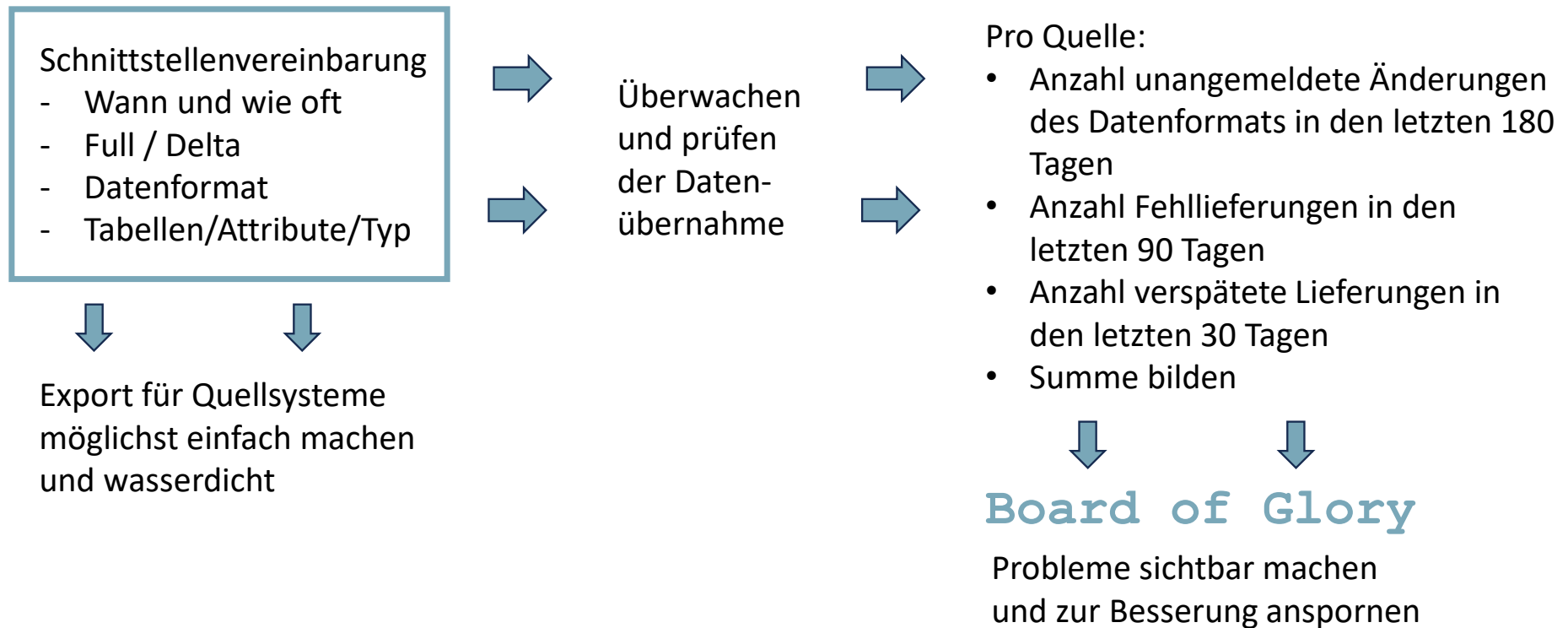


Besser sind
einfache Lösungen



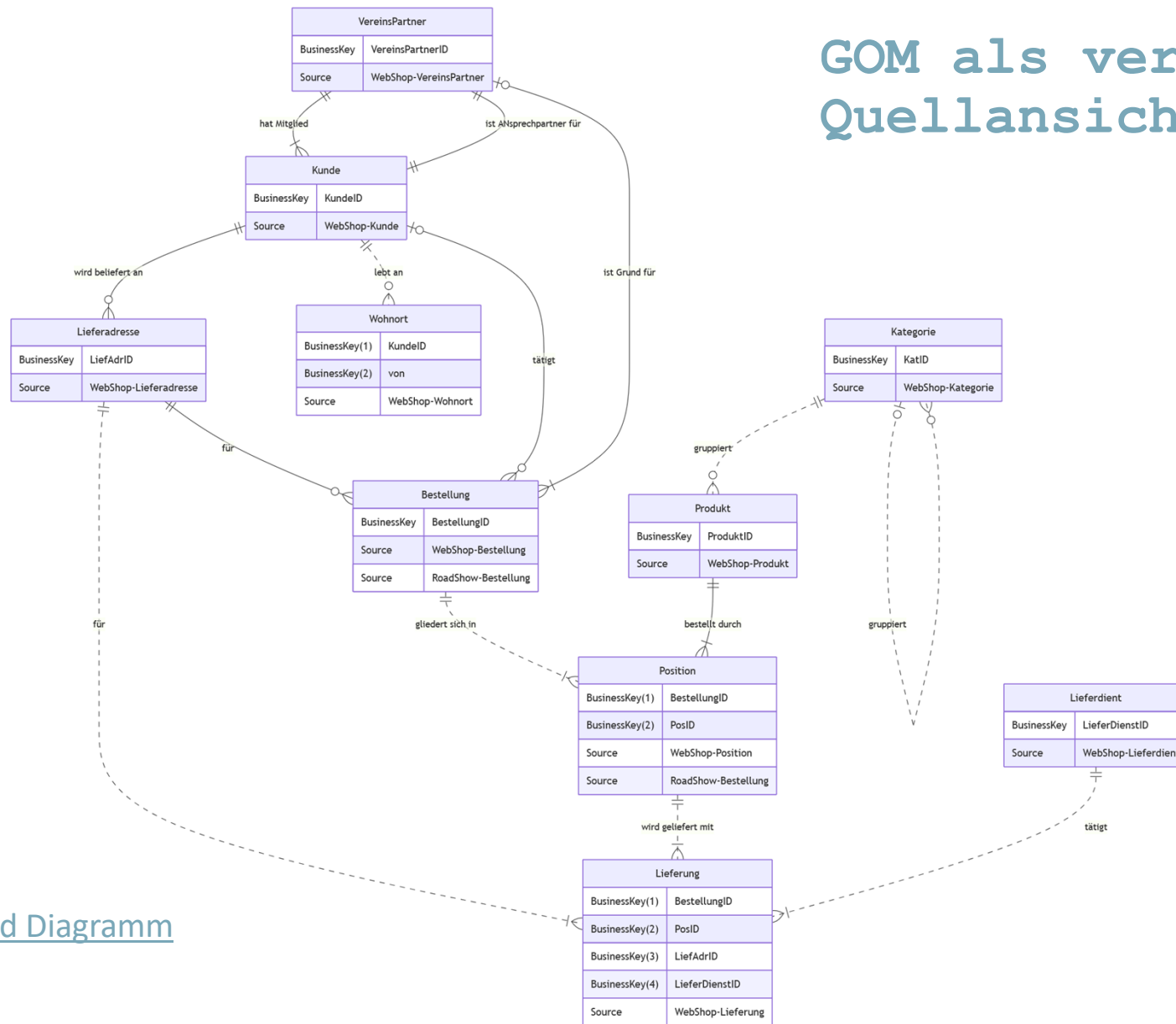
Standardwege, die die
Umsetzung operativ einfach
halten

Aktives Nutzen der Schnittstellenvereinbarung



GOM als verknüpfte Quellansicht

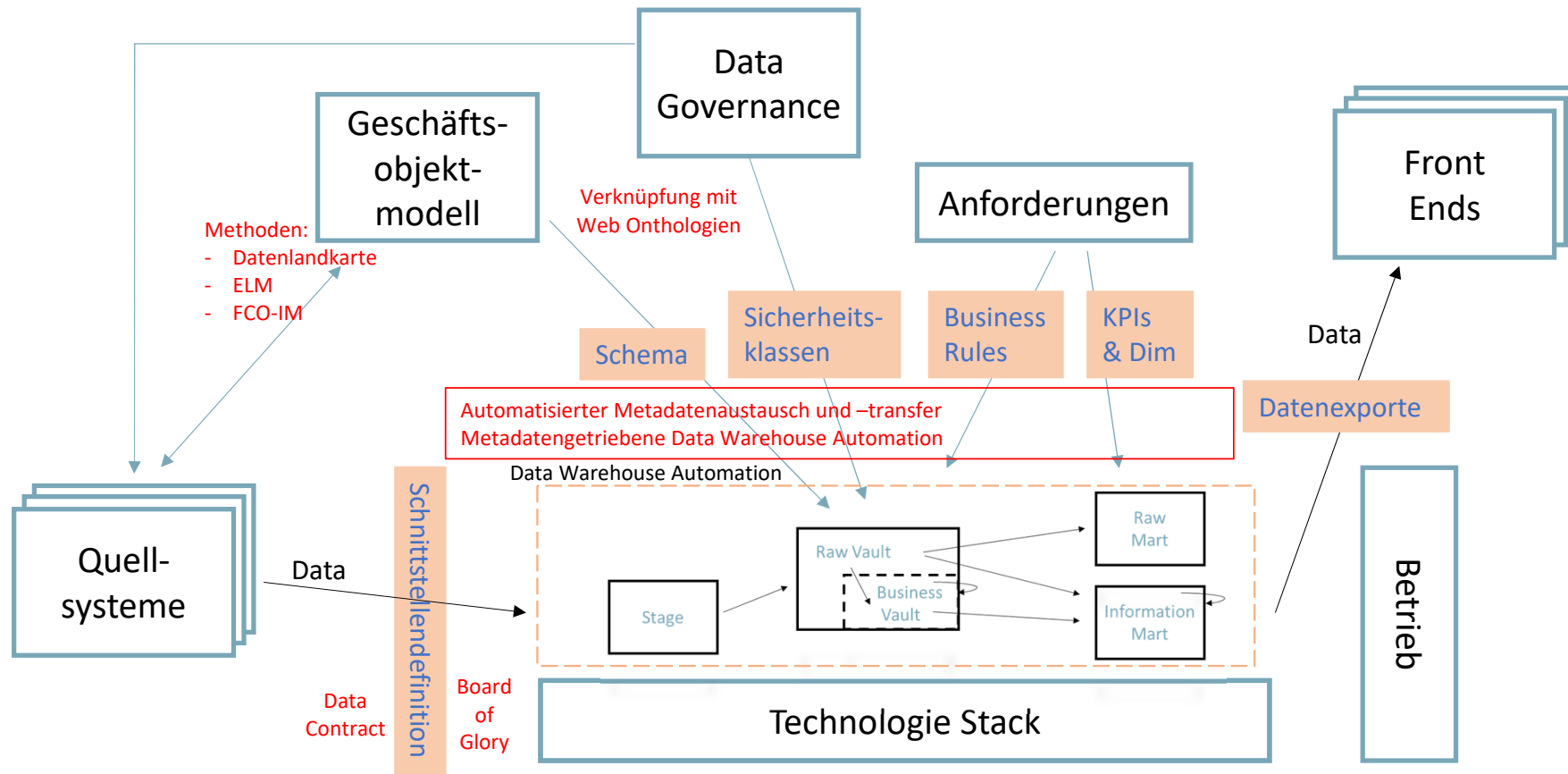
Willibald Business Object Source View



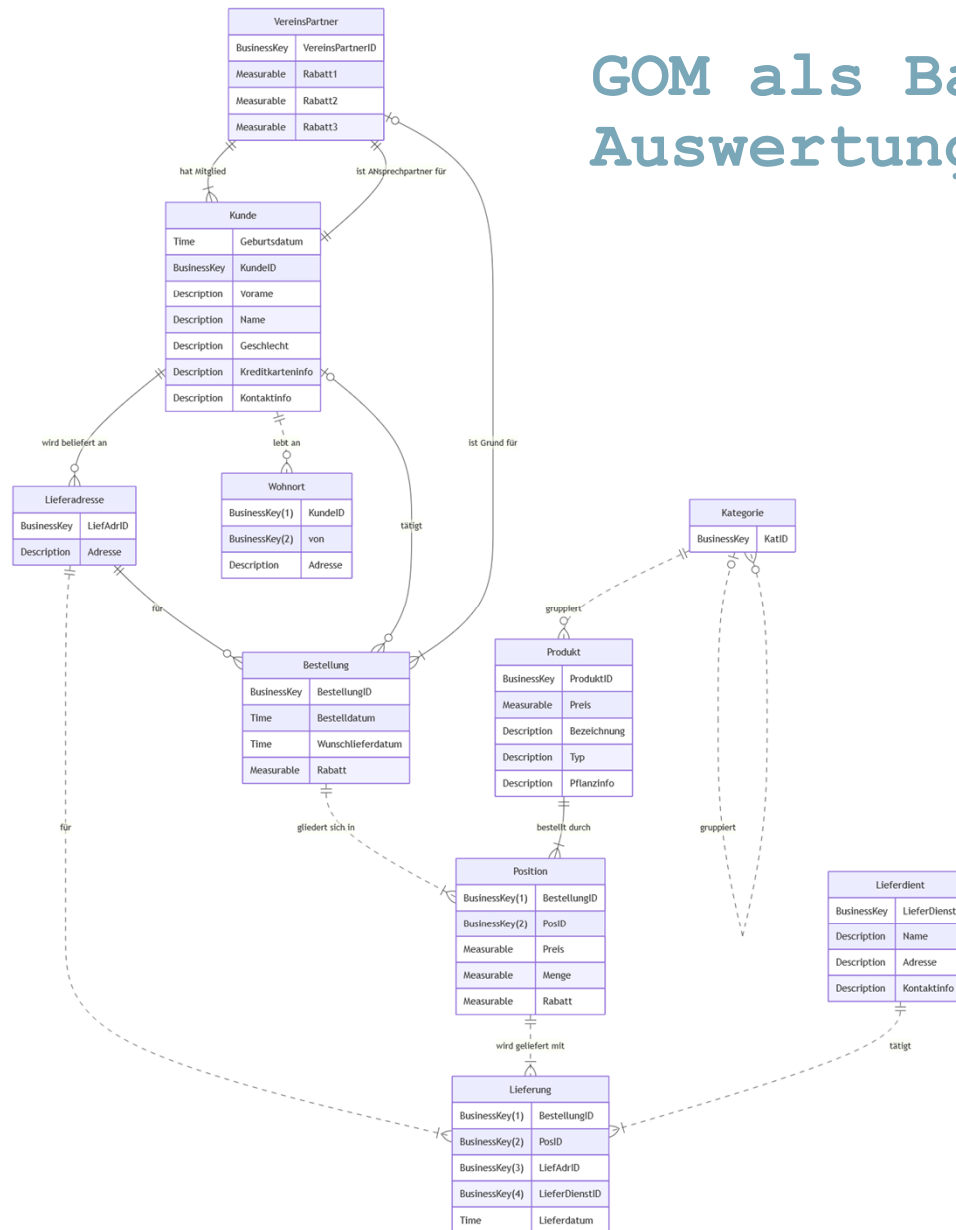
Mermaid Diagramm

Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

Datenqualität als gesonderten Prozess betrachten

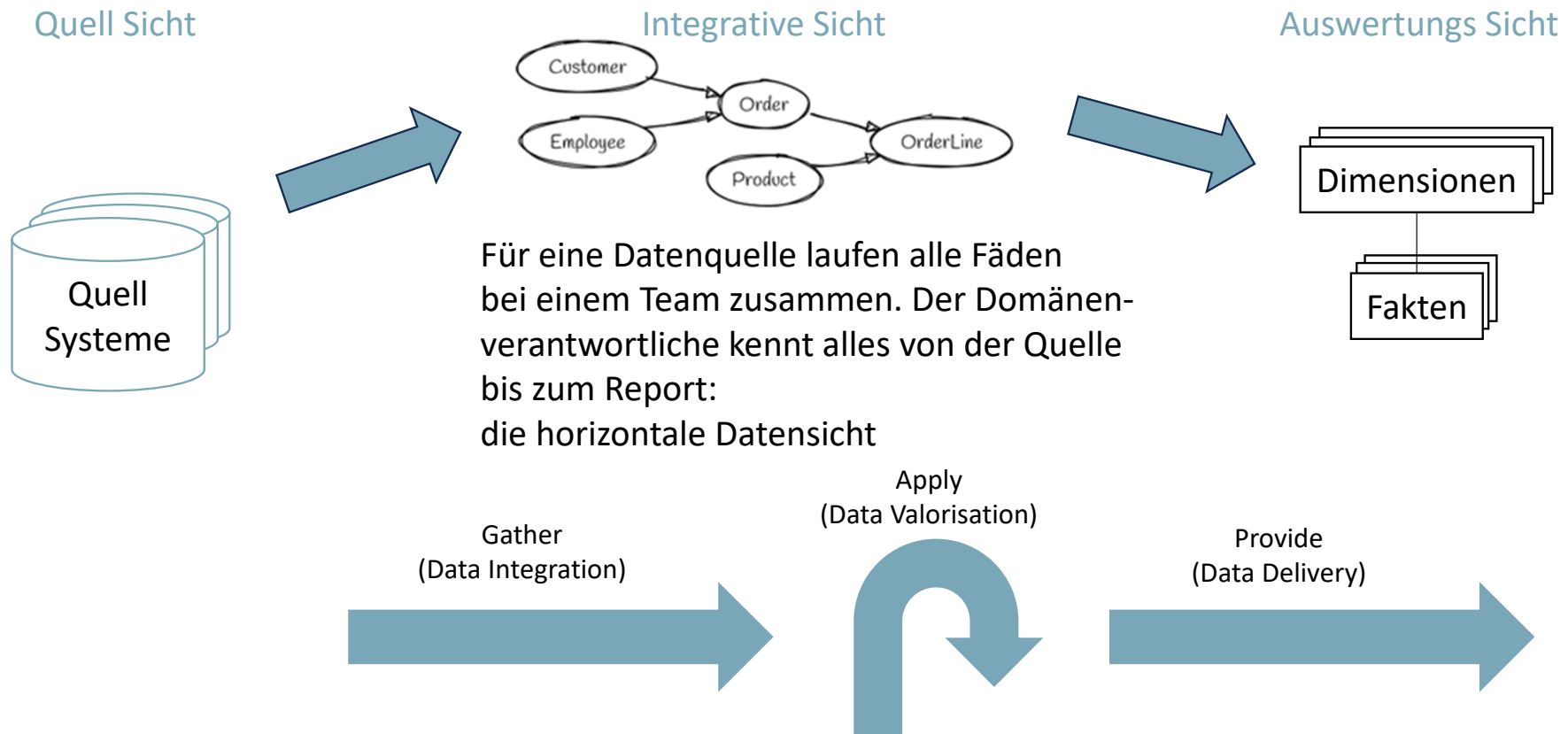


GOM als Basis für die Auswertungen



Mermaid Diagramm

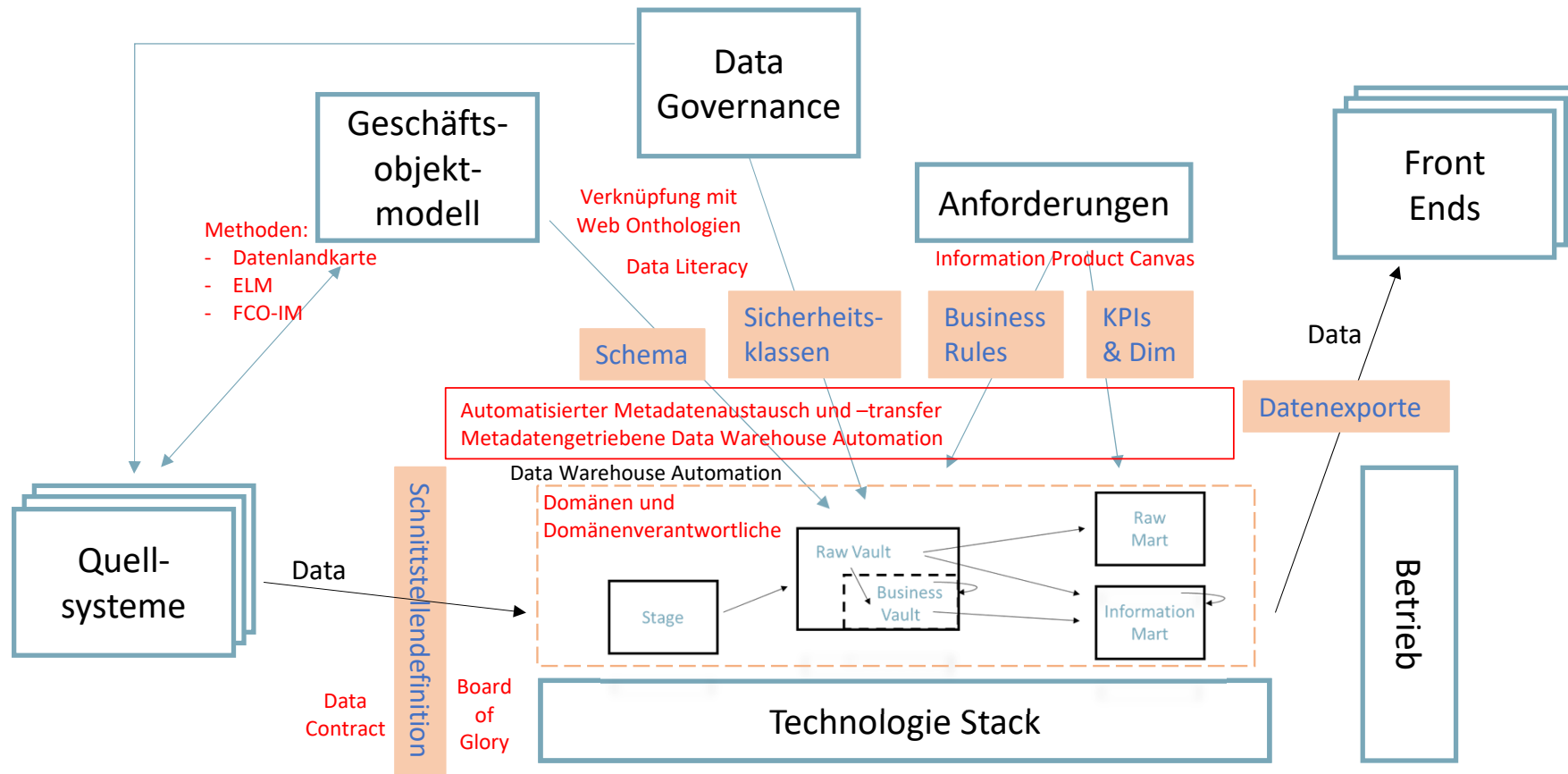
Domänenverantwortliche behalten den Überblick



Für eine Datenquelle laufen alle Fäden bei einem Team zusammen. Der Domänenverantwortliche kennt alles von der Quelle bis zum Report: die horizontale Datensicht

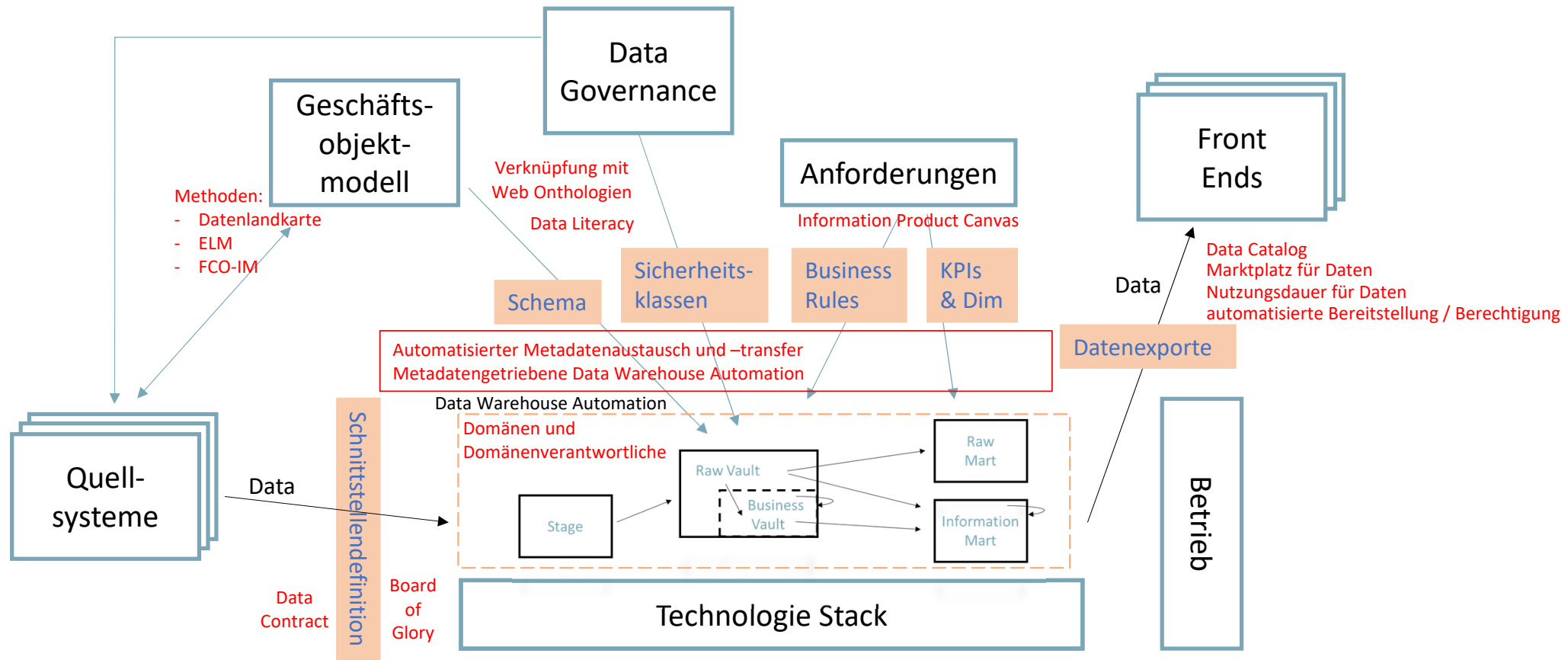
Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

Datenqualität als gesonderten Prozess betrachten



Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

Datenqualität als gesonderten Prozess betrachten



Einfaches Lineage mit minimaler Dokumentation

Zu jeder Prozedur im Business Vault werden Routine und Quelle bzw. Ziel in der Datenbank hinterlegt.

```
CREATE TABLE [dbo].[BizVaultMap] (  
    [Input] varchar(256) NOT NULL,  
    [TypeIn] char NOT NULL,  
    [Output] varchar(256) NOT NULL,  
    [TypeOut] char NOT NULL  
)
```

Mit den folgenden Werten
für TypeIn/TypeOut:

T	Tabelle
V	View
F	Funktion

Während des Deployment einfügen:

```
IF NOT EXISTS (SELECT 1 FROM [BizVaultMap]  
    WHERE [Input] = ,SAT_OrderItem'  
    AND [TypeIn] = 'T'  
    AND [Output] = ,P_Clean_OrderItem'  
    AND [TypeOut] = ,F')  
  
BEGIN  
    INSERT INTO [BizVaultMap] ([Input], [TypeIn], [Output], [TypeOut])  
    VALUES (' SAT_OrderItem', 'T', 'V_SR_CONFIRM_REMDOC_', 'V')  
END
```

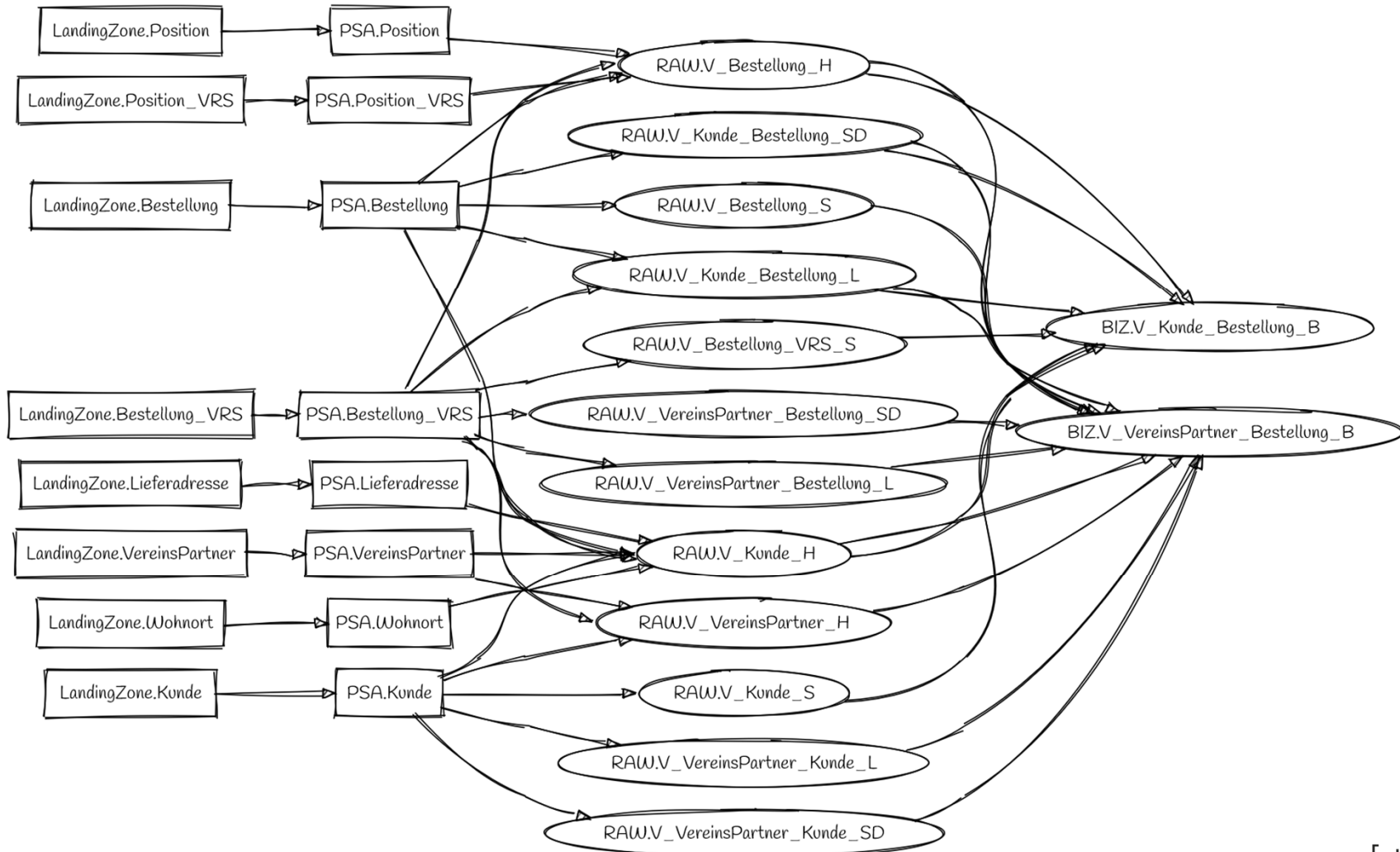
Die Tabelle liefert den Input für graphviz

```
select cvs
from (
  select 'b' reihe, (''' + [Input] + '"->"' + [Output] + ''') as cvs
  from [dbo].[BizVaultMap]
  union all
  select 'a' reihe, ''' + [Input] + '" ' +
    CASE WHEN [TypeIn] = 'T' THEN '[shape=box]'
    WHEN [TypeIn] = 'V' THEN '[shape=ellipse]'
    WHEN [TypeIn] = 'F' THEN '[shape=octagon]'
    ELSE '[shape=star]' END
  from [dbo].[BizVaultMap]
  union all
  select 'a' reihe, ''' + [Output] + '" ' +
    CASE WHEN [Typeout] = 'T' THEN '[shape=box]'
    WHEN [Typeout] = 'V' THEN '[shape=ellipse]'
    WHEN [Typeout] = 'F' THEN '[shape=octagon]'
    ELSE '[shape=star]' END
  from [dbo].[BizVaultMap]
) x
group by reihe, cvs
order by reihe
;
```

Selektion aller Vorgänger per rekursiver cte

```
with mapping as (  
    select *  
    from [Meta].[Lineage] a  
    where Output in ('BIZ.V_VereinsPartner_Bestellung_B', 'BIZ.V_Kunde_Bestellung_B')  
    union all  
    select a.*  
    from [Meta].[Lineage] a  
    inner join mapping m on m.Input = a.Output  
)
```

Und hier das Ergebnis

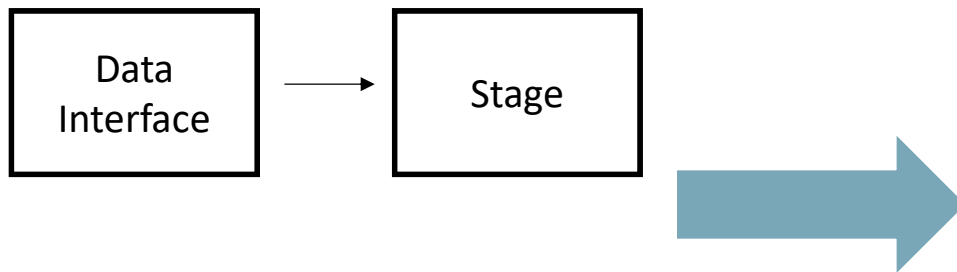




Was ist das Problem mit den Fehlern?

- Error Handling –
Wahrnehmung und Handling
- Prinzip der Sparsamkeit
 - Wer immer ‚Wolf!‘ ruft...
 - Fehler genau einmal melden
 - Behebung überwachen
 - Warnungen eher lassen
- Fehlerprüfung immer als optionales Feature

Error Handling im DWH – Meldungen dokumentieren



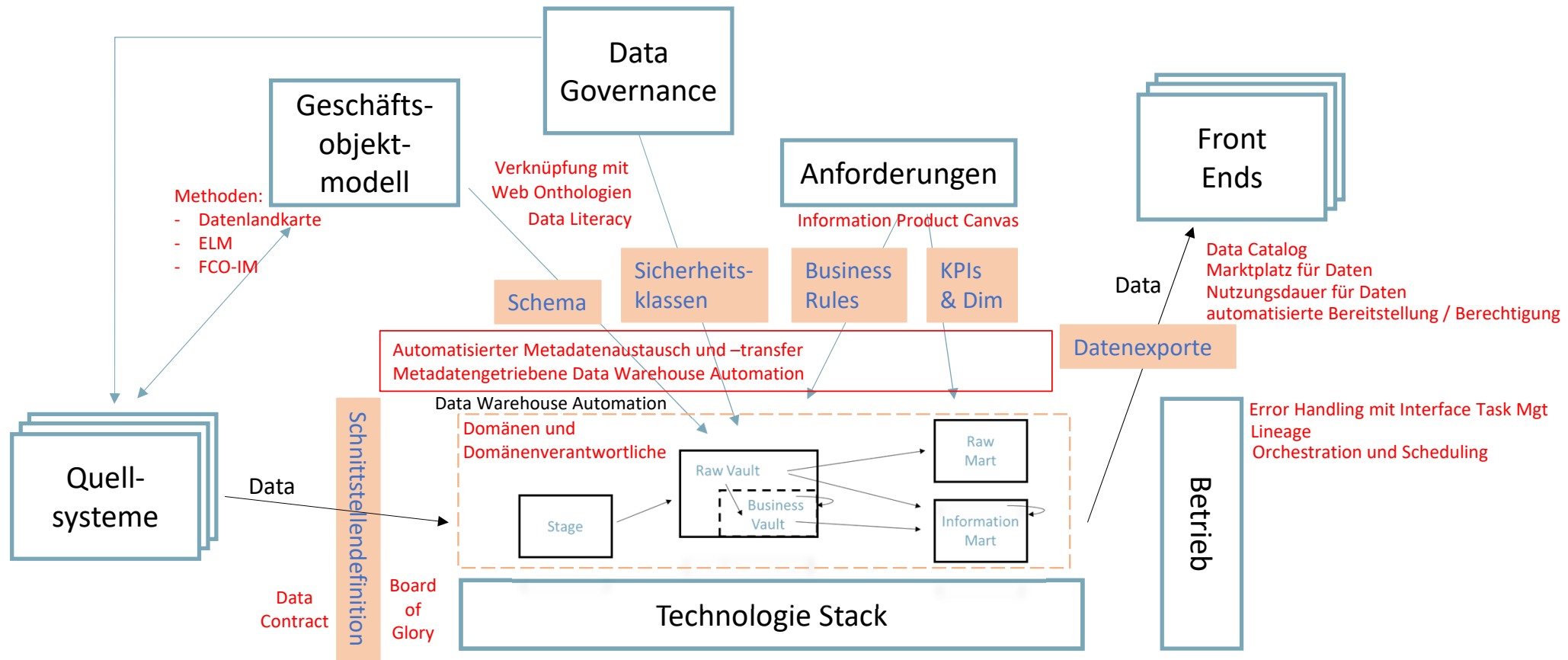
Schema-Änderung
Quelle (falls möglich),
fehlende Daten
(primary keys)

Dubletten,
Hard Rule Fehler

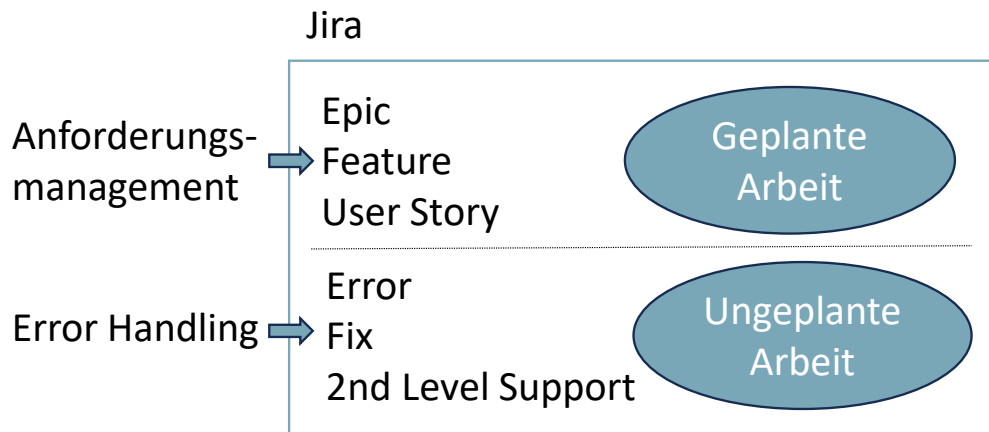
- ★ Meldungen definieren und dokumentieren
- ★ Verantwortlichen definieren
- ★ mögliche Meldungen:
 - ★ Dubletten
Ein Primärschlüssel wird in einer Lieferung einer Tabelle mehrfach geliefert.
 - ★ Fehlender Schlüsselbegriff
Der Primärschlüssel fehlt. Diese Meldung wird pro Tabelle als Gesamtzahl ausgegeben.

Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

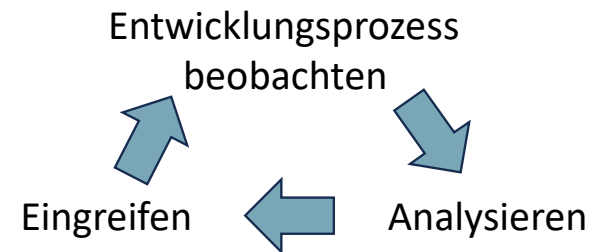
Datenqualität als gesonderten Prozess betrachten



Ungeplante Arbeit tracken



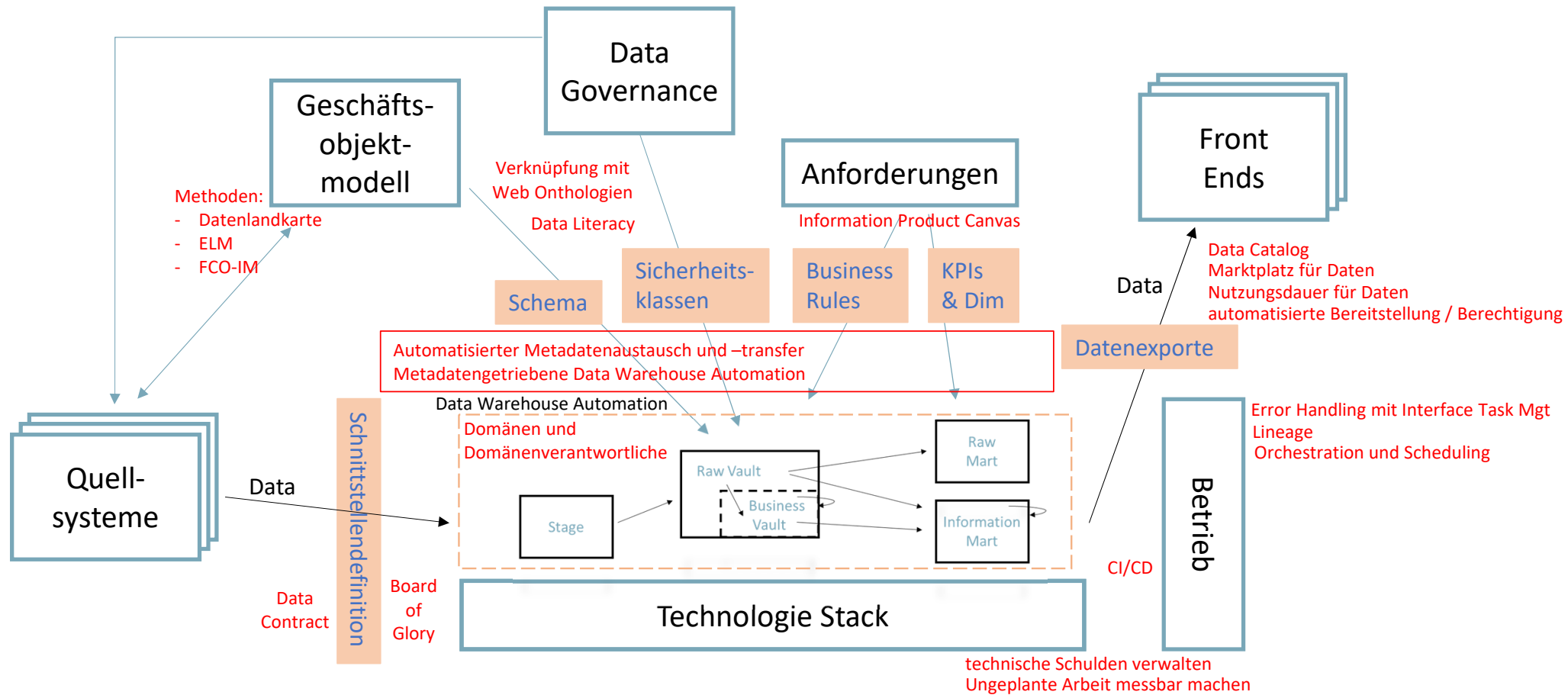
- Ungeplante Arbeit ist ein Hindernis für schnelle Entwicklung



- Tracking der Aufwände liefert wichtige Fakten zur Verbesserung

Konzepte für bessere Zusammenarbeit bzw. für ein besseres Data Warehouse

Datenqualität als gesonderten Prozess betrachten





Ist das nicht
zuviel?

- ★ Die fertige Lösung für ein ideales DWH gibt es noch nicht.
- ★ Jede der hier beschriebenen Lösungen hilft auch für sich allein
- ★ Gleichzeitig ist das Big Picture klar, Sackgassen Lösungen lassen sich vermeiden
- ★ 15 % gehen immer

Danke für Ihre Aufmerksamkeit!

m[method] 2 data