



# Zwischen Skylla und Charybdis

## Automatische vs. manuelle Content-Migration eines großen Informationsportals

# Agenda

- Einleitung: Migration unter der Oberfläche
- Sorgenkind Migration
- Vorgehen im Projekt
- Probleme
- Fazit



# Sorgenkind Migration: Dokumentation Altsystem

***(F1) Erklären Sie sich bereit, die Migration auch durchzuführen, wenn Sie keine zusätzlichen Informationen zum aktuellen Softwaresystem erhalten. Erklären Sie bitte hierzu Ihre Bereitschaft. (J/N)***

**Ja .**

- Keine Zugriffsschicht, keine Alt-Software
- Keine (brauchbare) Dokumentation
- Keine Auskunft
- Input: Datenbank-Dump und laufendes Portal
- Fester Zieltermin (Abschaltung Altsystem)

## Sorgenkind Migration (II): Kommunikation & Planung

- Kunde will zeitnahe Informationen, kann aber abstrakten, software-technischen Vorgang der Daten-Migration oft nicht einschätzen
- Ideale Welt: Migration nach Entwicklung des Zielsystems
- Realität: Parallelisierung mit „Henne – Ei Problem“
  - Ziel-Datenmodell ist Voraussetzung für Portalsoftware und Migration
  - Designer wollen Infos über Altsystem von Migration für Datenmodell
  - Migration will finales Modell und Software zur QS der Ergebnisse
  - Portalsoftware stellt neue Anforderungen an das Datenmodell
  - Entwicklung will Migrationsergebnisse als Testdaten
- Vorteil: Frühzeitige Echtdateen für QS der Software

## Vorgehen im Projekt (I): Planung

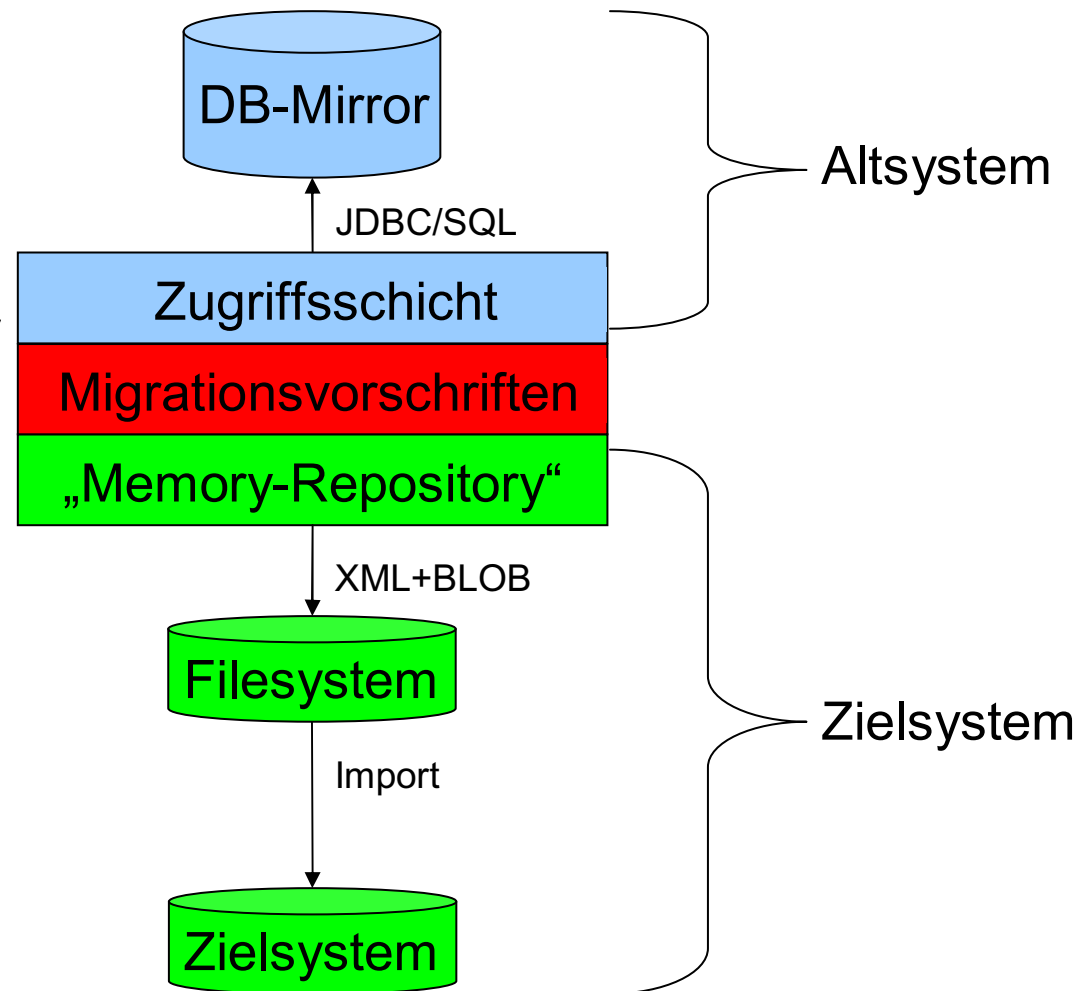
- Weder voll-automatische noch voll-manuelle Migration möglich
  - Große Mengen automatisch migrieren
  - Einzelfälle und Feinschliff manuell
- Kein Big Bang → Iterationen
- Code-Freeze Altsystem
- Große Migration, QS und Nacharbeiten (Parallelbetrieb), Delta-Migration
- 1 Person für technische Migration
- Ab Projektmitte 1 Person als fachliche Schnittstelle für Migration

## Vorgehen im Projekt (II): Technische Strategie

- Zielsystem hat geeignete XML-Import-Schnittstelle
- Frage ist Extraktion der Daten aus Altsystem nach XML-Vorgabe:
  - DWH / ETL verworfen wegen Lizenzkosten
  - O/R Mapper verworfen wegen DB-Schema
  - XML-Dump & XSLT verworfen wegen Performance & Mächtigkeit

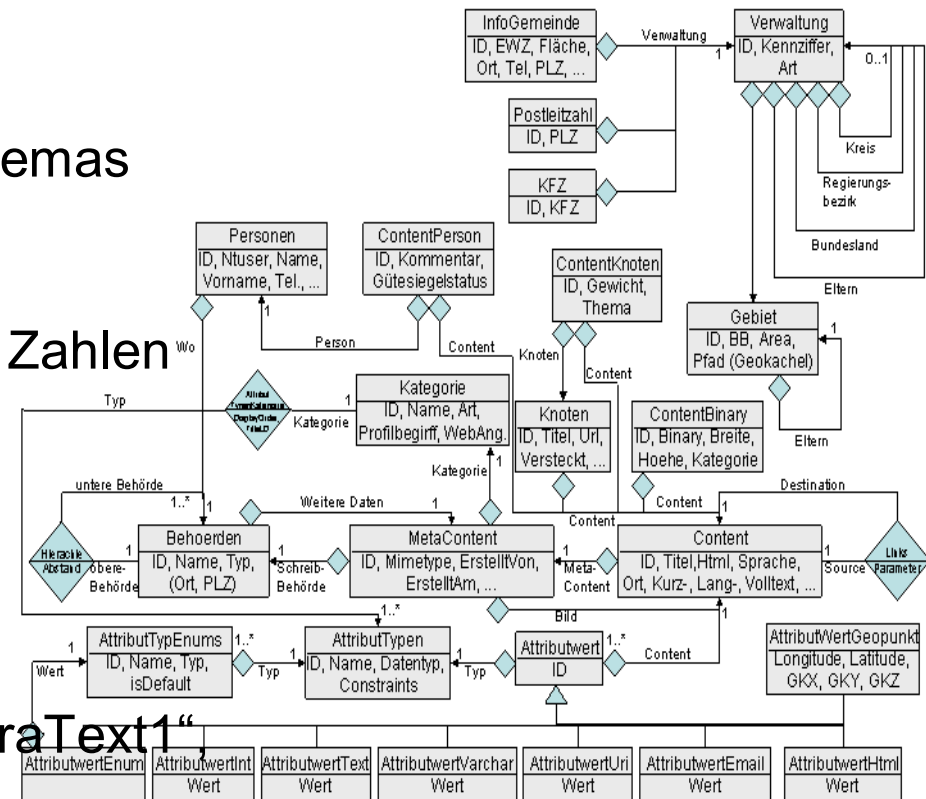
# Vorgehen im Projekt (III): Migrationsarchitektur

- Brute-Force mit Java
- Eigene, minimalistische Zugriffsschicht
- JTidy zur HTML-Reparatur
- Migrationsvorschriften als Wegwerfcode
- BLOBs direkt auf Platte
- Content komplett im Hauptspeicher als Repository
- Daten per Import ins Zielsystem



# Probleme (I): Ausgangsschema

- Aufwendiges Reverse-Engineering
- Geringe Qualität des Datenbankschemas
  - Keine Primärschlüssel
  - keine Referenzen, nur identische Zahlen
  - „Tote Links“
  - Völlig de-normalisiertes Schema
  - Jeder Datensatz hat 2 „IDs“
  - DB-Felder wie „volltextDirty“, „extraText1“, „extraInt1“, „mailversandZaehler“, ...
- Viele implizite Regeln in der Altsoftware



## Probleme (II): Struktur & Verlinkungen

- Inhalte:
  - im Zielsystem in Ordnerstruktur
  - im Quellsystem als Trefferliste.
- Navigationsziele als URL-Strings mit codierten Parametern
- „Code“ im Content (`<!-- IfLoggedIn content="profile:ausschreibung" -->...`)
- HTML-Fragmente mit beliebiger Syntax
  - `<v:shape id=_x0000_i1025 style="WIDTH: 12.75pt; HEIGHT: 11.25pt" o:ole="" type="#_x0000_t75">  
<v:imagedata o:title=""  
src="file:///C:/DOKUME~1/mustermann/LOCALS~1/Temp/msoclip1/03/clip_image001.png"><FONT  
size=2></FONT></v:imagedata></v:shape>`
  - `<DIV3><TABLE border=1 <tr summary>Title</TABLE></div3>`
- Interne Links als externe URLs (`www.bund.de/GZW/Natur-und-Umwelt-.4489.htm?view=special`)

## Probleme (III): Zielsystem

- XML-Import der Daten dauerte fast zwei Tage (Export ca. 30 Min.) → wenig Zeit für Tests / Iterationen
- Automatische Publikation führte zu OutOfMemoryError (bei 1.8 GB Heap)
- Fehler in CMS-Standardsoftware führte bei Publikation zu inkonsistentem Zustand der Live-Daten

# Fazit

- Migration ist mehr als das Verschieben von Content
- Datenmigration ist einmalige Sache → Wegwerfsoftware → muss nicht schön sein
- Trennlinie zwischen automatischer und manueller Migration finden
- Parallelisierung von Software-Entwicklung und Migration vermeiden – aber oft nicht zu vermeiden
- Gute Kommunikation mit dem Kunden wichtig, da das Thema für ihn fachfremd und nur schwer überschaubar, ihn aber direkt betrifft
- Klare und einfache Spielregeln für alle Beteiligten aufstellen:
  - Was ist erlaubt? Was ist verboten?

# Fragen?



Migration?  
➔

