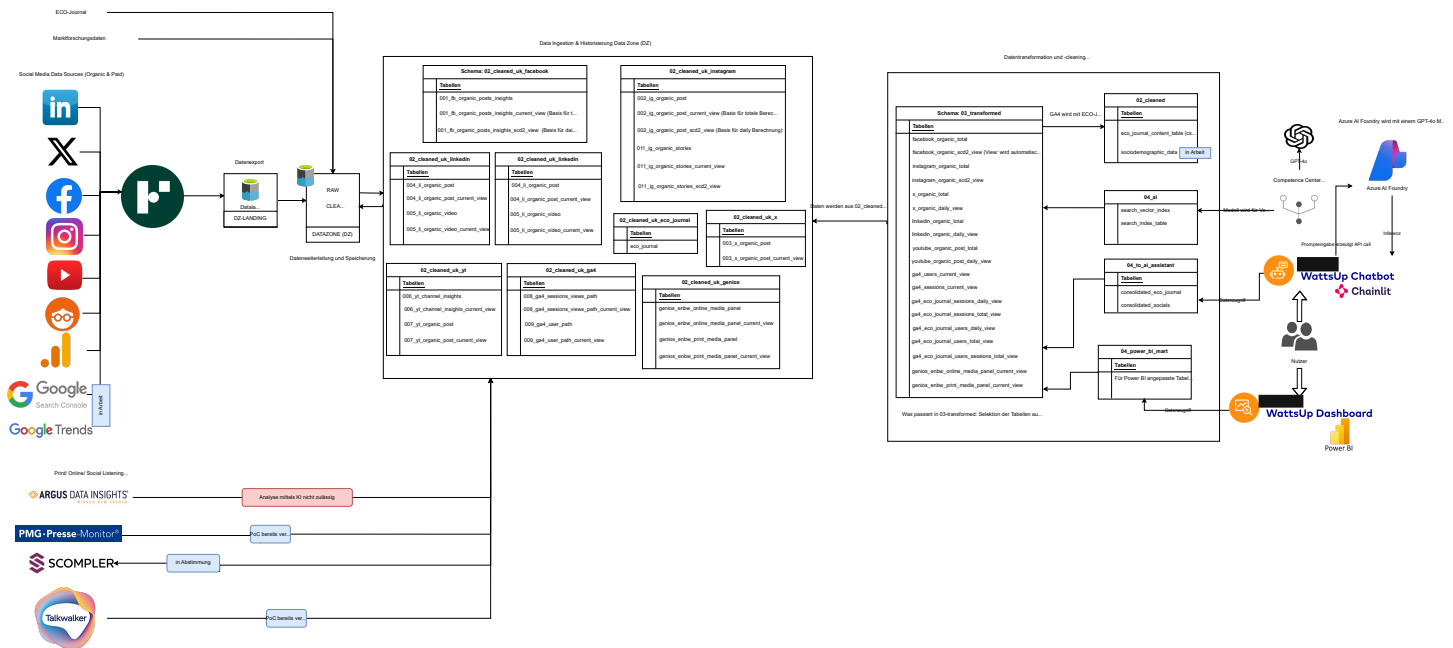


UK übergreifende Architektur

Last updated by | [REDACTED] | 15 Oct 2025 at 16:30 CEST

Diese Architektur zeigt eine modulare Datenplattform, die verschiedene Social-Media- und Content-Datenquellen verarbeitet, bereinigt, historisiert und sowohl für Reporting als auch für KI-gestützte Analysen bereitstellt. Die Integration von GPT-4o zur automatisierten Textanalyse (Abstracts & Tags) und die Anbindung an Power BI ermöglichen datengetriebene Entscheidungen und Kommunikation.



Link zur Architekturskizze [hier](#).

1. Datenquellen

Es werden Daten aus diversen Quellen bezogen:

Social Media (Organic & Paid) via Funnel

- LinkedIn
- X (Twitter)
- Facebook
- Instagram
- YouTube
- TikTok
- Google Analytics / GA4 -> ECO-Journal basiert auf GA4 Daten
- Google Search Console (in Arbeit)
- Google Trends (in Arbeit)

Print / Online / Social Listening

- Argus Data Insights – Analyse mittels KI derzeit nicht zulässig
- Genios (On hold)
- Scompler – in Abstimmung

Weitere Quellen

- ECO-Journal
 - Marktforschungsdaten (in Arbeit)
 - Soziodemografische Daten (in Arbeit)
-

2. Datenexport und -speicherung

- Die Daten werden von Funnel in DZ-LANDING gepusht. Dies dient als extra Sicherheitsschicht, wenn ein externes Quellsystem ins DaTIF Daten pusht.
 - Daten werden in die DZ mittels Copy Activity kopiert und dort prozessiert und historisiert. Die Rohdaten werden in 01-raw gespeichert. Die historisierte Tabellen werden als Delta Table in 02-cleaned in der DZ persistiert.
-

3. Data Ingestion & Historisierung (Data Zone - DZ)

Strukturierte Speicherung in verschiedenen Schemas und Tabellen. Es gibt pro Quellsystem ein gesondertes Schema. Dies hat einen Governance-rechtlichen Hintergrund, sodass Fachbereiche dedizierten Zugang zu ausgewählten Daten erhalten können.

- 02_cleaned_uk_facebook (z.B. `organic_post_insights`): Bei den Meta Kanälen bilden die Lifetime Values die Summe einzelner Metriken für den gesamten Verlauf eines Posts ab. Die Daily Änderungen werden durch die SCD2-View ermittelt.
 - 02_cleaned_uk_instagram (z.B. `stories`, `posts_current_view`): Lifetime Values und die daily Werte werden über SCD2 View ermittelt.
 - Bei den nachfolgenden Datenquellen werden täglich die Änderungen exportiert. D.h. die current View stellt die Daily View dar. Die totale Aggregation werden nachträglich in der PZ Uk berechnet.
 - 02_cleaned_uk_linkedin:
 - 02_cleaned_uk_x (Twitter)
 - 02_cleaned_uk_yt (YouTube)
 - 02_cleaned_uk_ga4 (Google Analytics 4)
 - 02_cleaned_uk_eco_journal
 - 02_cleaned_uk_genios (Print/Online Media Panels)
-

4. Datentransformation & -Cleaning (Product Zone - PZ UK)

a) 03_transformed

In der Product Zone (03_transformed) werden Daten bereinigt, zusammengeführt und in logisch aufbereitete Tabellen überführt, u.a.:

- linkedin_organic_total: Konsolidierung der beiden Tabellen Posts und Videos
- facebook_organic_total
- instagram_organic_total
- x_organic_daily_view
- youtube_organic_post_total
- ga4_sessions_current_view, usw.
- ECO-Journal-Sessions und -User-Daten
- noch genauer ergänzen [REDACTED]

b) 04_to_ai_assistant

- Konsolidierung der Tabellen, welche dem Azure OpenAI Client zur Verfügung gestellt werden bzw. dort hochgeladen werden. Neue Architektur, sodass wir csv's über die AzureOpen AI Assistant API direkt dem Client zur Verfügung stellen. Die Tabellen werden als csv's hochgeladen und können vom Chatbot durchforstet werden. Folgende Tabellen werden bisher in 04_to_ai_assistant konsolidiert:
- Konsolidierte Datenquellen als gesonderte Tabelle thematisch gruppiert:
 - consolidated_eco_journal
 - consolidated_socials: Enthält alle Social Media Daten. * Attribute bitte noch nachpflegen

c) 04_to_ai

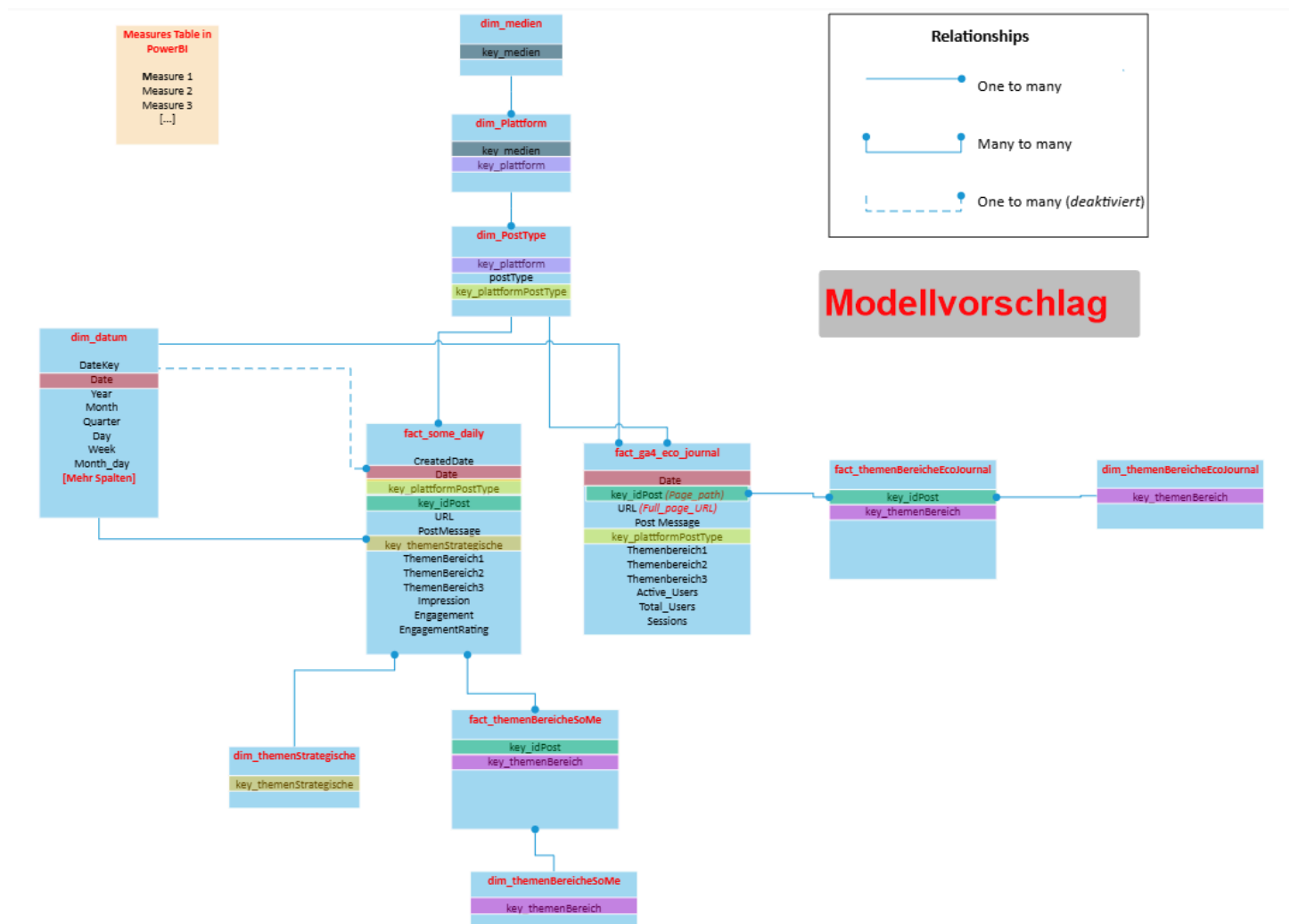
- Erstellung von Vektorindex sowie Search Index via Databricks. Diese werden für die semantische Vektorsuche verwendet.

d) CC GenAI für Datenanreicherung

- Das GPT-4o Modell vom CC GenAI der [REDACTED] wird für die Abstract Generierung, Tagging von Tonalität sowie der strategischen Themen sowie Themenbereiche verwendet. CC Gen AI stellt einen API Endpoint für deren gehostete Modelle zur Verfügung.

e) 04_power_bi_mart

- Dies ist der Modellvorschlag (muss noch bestätigt werden). Alle Tabellen werden in Databricks und Maßnahmen in PowerBI erstellt.



5. Nutzerinteraktion

Endnutzer greifen auf unsere Daten durch folgende zwei Systeme zu:

a) WattsUp Chatbot

- Zugriff auf gepushte csv's welche in 04_to_ai_assistant liegen.
- Verwendung des Chainlit Frameworks

b) WattsUp Power BI Dashoard

- Zugriff auf Tabellen aus 04_power_bi_mart