

Mobile Erfassung von Geofachdaten

Eine service-orientierte Betrachtung

Bastian Ellmenreich (UM), Clemens Döpmeier (KIT-IAI)

Institut für Automation und angewandte Informatik (IAI)



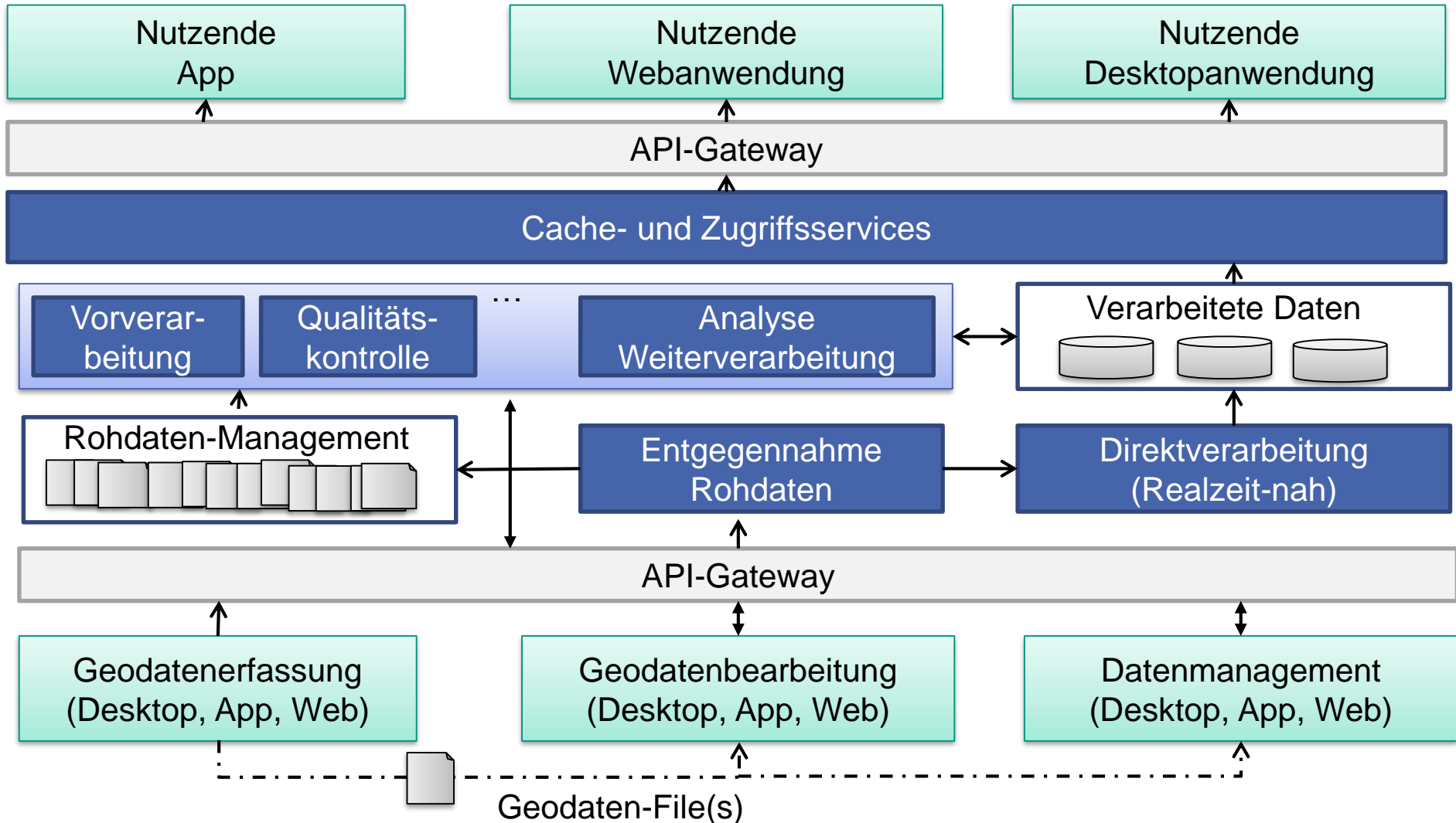
- Motivation
- Ein anwendungsübergreifender, service-orientierter Ansatz zur Geodatenerfassung, Management, Verarbeitung und Bereitstellung
 - Übersicht
 - (Mobile) Erfassung
 - Datenmanagement
 - Qualitätskontrolle, Datenbearbeitung, Anreicherung und Analyse
 - Datenbereitstellung
- Zusammenfassung

Motivation

- Frontends zur mobilen Datenerfassung (Apps) stellen nur **einen kleinen Teil** des notwendigen **Ökosystems zur (Geo)datenerfassung** dar
- Die **Schnittstellen** zu Ihnen werden wesentlich durch das „Backend“-Ökosystem zur Weiterverarbeitung bestimmt
- Dieses muss komplexe Aufgaben übernehmen
 - Daten entgegennehmen und transformieren, anreichern, ...
 - Speicherung der Daten
 - Möglichkeiten zur Qualitätskontrolle vor Nutzung bieten
 - Für schnellen Zugriff Daten in Caches überführen
- Die **Offenheit** und **Modularität** des Backend-Systems bestimmt wesentlich die Schnittstellen zu den mobilen Erfassungsanwendungen
 - Bzgl. **API-Schnittstellen + Datenformaten**
 - Nutzung von **Standards** ist hier sehr **hilfreich**
- Ein **service-orientiertes Gesamtkonzept** ist ideal für den Aufbau eines solchen Ökosystems unter Nutzung von Standards und verfügbaren Werkzeugen

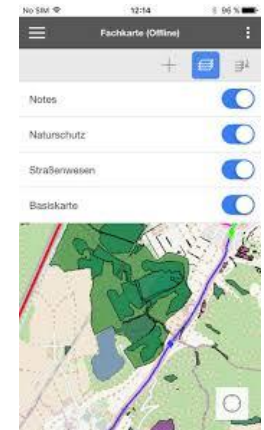


Anwendungsübergreifende Geodaten-Erfassung, Verarbeitung und Bereitstellung



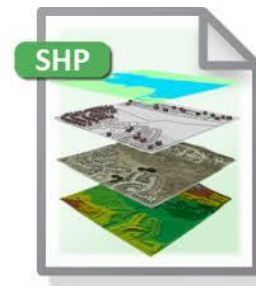
Geodatenerfassung

- Verschiedene Formen
 - Mit Spezialgeräten
 - Mobilgeräten und Apps



- Weiterleitung
 - Per API an Service (Bsp.: OSM Editier-API)
 - Als File an Anwendung zur Vorverarbeitung und anschließendem Upload (über API) oder Web-Interface
- Viele unterschiedliche standardisierte Formate verfügbar und nutzbar

- JSON oder XML-basiert (GeoJSON, KML, GML, GPX)
- Binärformate (Shapefile, GeoTIFF)
- Anwendungsspezifisch oder generisch



- Je standardisierter die API-Schnittstellen (z.B. OpenGIS APIs, OSM APIs) und Formate, desto größer die Nutzbarkeit generischer Dritt-Anwendungen

Mobile Beispielanwendungen

■ Erfassung von Geogrunddaten

- Z.B. Zenon Leica GG04 Plus
 - GPS-Erfassungsgerät mit Korrektur im cm-Bereich + Tablet-basierte APP zur Datenaufnahme und Weiterleitung
- Drohnen-basierte Erfassung ist im Kommen, insbesondere zur 3D-Erfassung



Zenon GG04 Plus von
Leica Geosystems

■ Generische Smartphone-Apps

- Outdoor-Navigationsanwendung
 - mit Tracking und Punkterfassung
 - OSM-Kartierungsapps (Bsp. OSMAnd) unterstützen OSM Editier-API

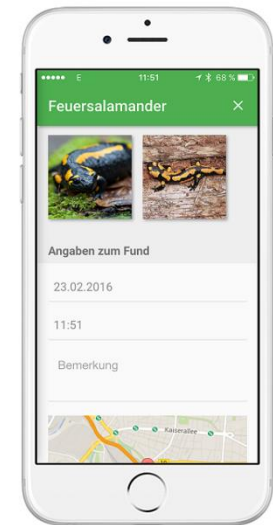


■ Erfassung spezieller (Umwelt)geodaten

- „Meine Umwelt“ für Meldung von Umweltbeeinträchtigungen
- Cadenza Mobile für Arbeit im Feld



OSMAnd



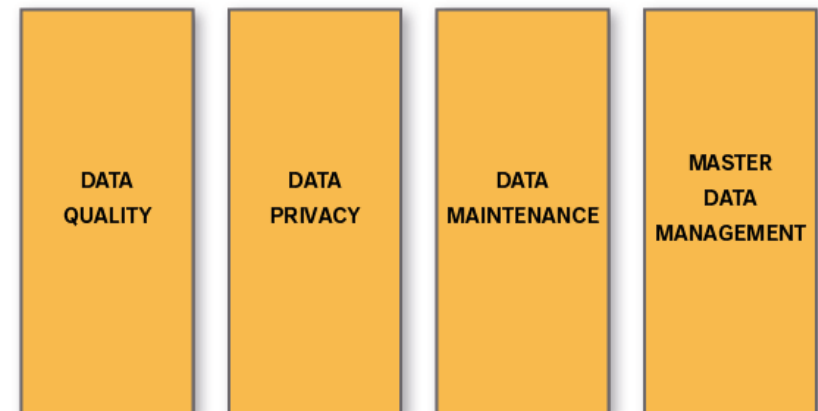
MeineUmwelt

(Roh)-Datenmanagement

- Entgegenkommene Daten sollten zunächst „as is“ gespeichert werden
 - Um Fehler in Rohdaten nachvollziehen zu können
 - Dafür eignen sich Datei-basierte Speicher mit Metadatenerfassung
 - Z.B. Verteilte Dateisysteme mit Objektmanagement (bei großen Mengen, z.B. Ceph oder HDFS)
 - Und/oder Asset-Managementsysteme on top
 - Integriertes Versionsmanagement
- Bestimmte Daten (geobasierte Meldungen) können zusätzlich in realzeit-nahe Weiterbearbeitung übergeben werden
- Im Zuge der Digitalisierung können service-orientierte Data Governance Anwendungen zum Datenmanagement geschaffen werden
 - Oft noch „zu Fuß“-Arbeit



VIER SÄULEN DES DATA GOVERNANCE



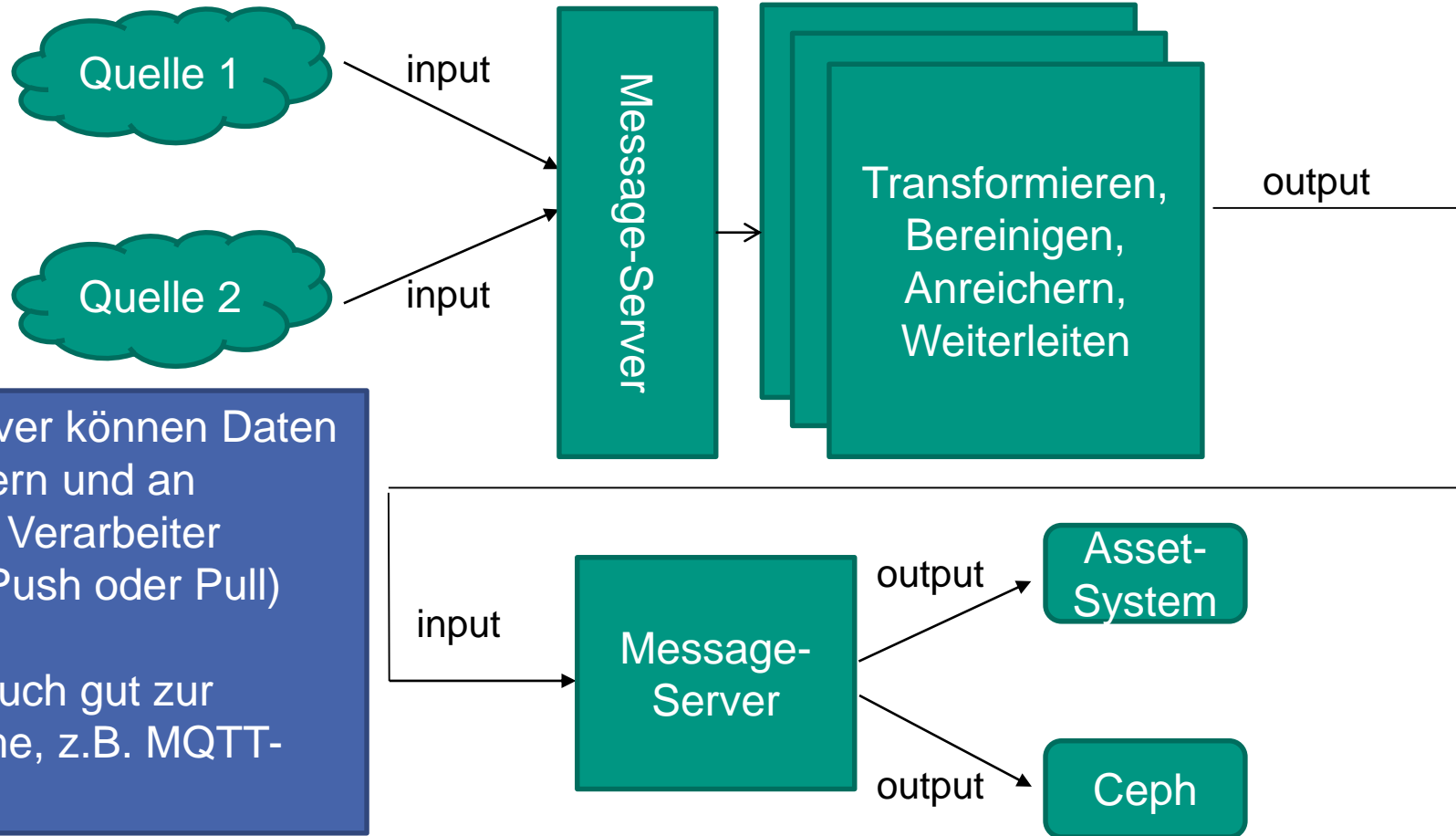
Entgegennahme, Weiterverarbeitung der Rohdaten

- Entgegennahme, Bereinigung, Transformation, Anreicherung
 - Kann durch dedizierte Anwendung passieren, aber
 - Es gibt eine Vielzahl von service-orientierten Workflow-Tools für die Entgegennahme und Weiterverarbeitung von (geo-basierten) Rohdaten
 - Lassen sich service-orientiert auch miteinander kombinieren
 - Bsp.: Logstash, Flume, NiFi
- Qualitätskontrolle
 - Dedizierte **interaktive** GIS-Anwendungen mit entsprechenden Schnittstellen
 - Web-/Mobil-/ oder Desktop
 - Schnittstelle zum Download zu kontrollierender Daten und zum Upload von Änderungen
 - Bsp.: [OSM Quality Assurance Editor](#) nutzt OSM Editier-API
 - Dedizierte Batch Data Analytics Anwendungen (OSM Lint), die Datenbestand auf grundlegende Fehler überprüfen
 - Machine Learning Anwendungen könnten hier auf Dauer komplexere Prüfungen übernehmen

Datenerfassung mit Frameworks, wie Flume, NiFi, Logstash

Ansatz mit Message-Server für Elastizität und Fehlertoleranz

- skalierbar
- flexibel



Message Server können Daten zwischenpuffern und an verschiedene Verarbeiter weiterleiten (Push oder Pull)

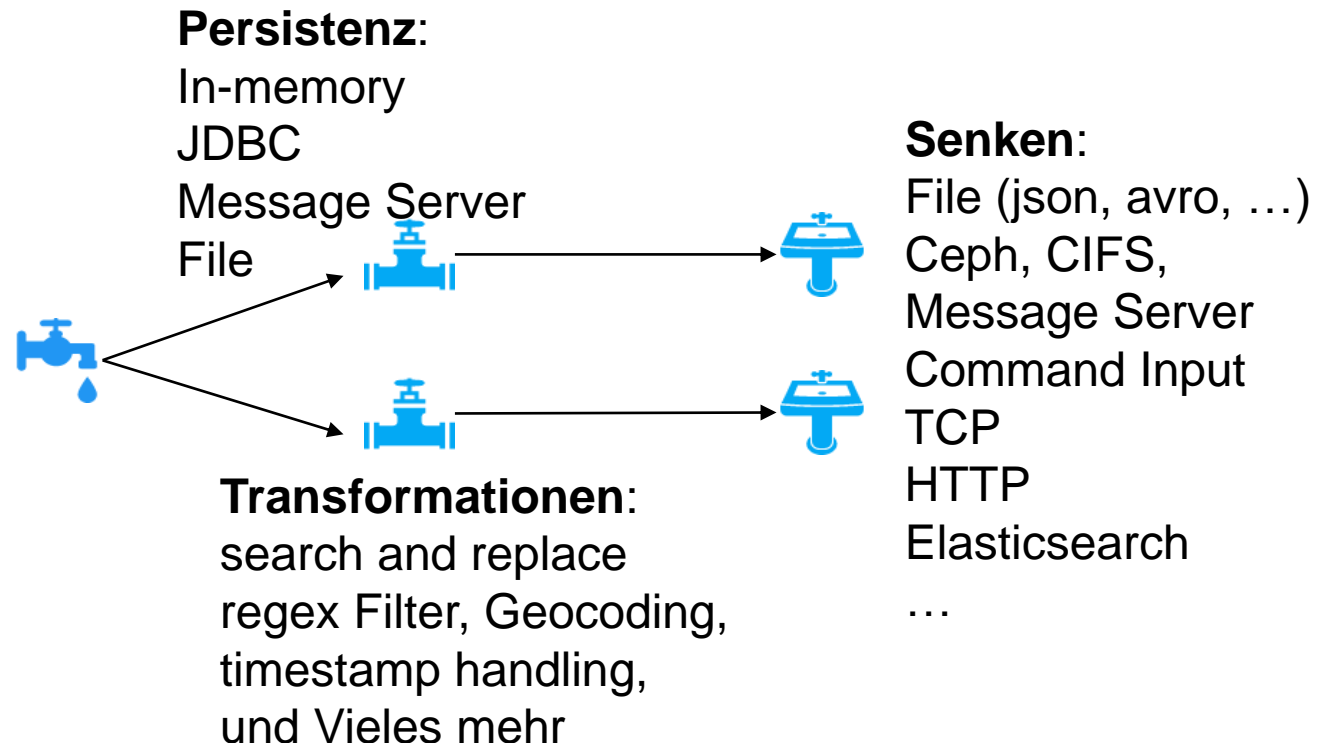
Eignen sich auch gut zur Datenannahme, z.B. MQTT-Server

Beispiel „Apache Flume“

- **Quellen** nehmen Daten entgegen
- Quellen unterstützen viele verschiedene Schnittstellen
- In **Kanälen** (Pipelines von Bearbeitungsschritten) werden die Daten transformiert, angereichert, bereinigt und zwischengespeichert
- Und dann an **Senke** geschickt, welche die Daten weiterleitet

Quellen:

File (json, xml, ...)
 Ceph, HDFS
 Message Server
 Command Output
 TCP
 HTTP
 ...



Elasticsearch und der ELK-Stack

- Der ELK-Softwarestack besteht aus 3 Open Source Projekten der Fa. “Elastic” (formerly Elasticsearch)
 - E => Elasticsearch (**Speicherung bearbeiteter Daten**)
(Hoch skalierbare NoSQL-Dokument-orientierte-Datenbank mit zusätzlichen Fähigkeiten für Analytics und Volltextindizierung)
 - Unterstützt auch Geodaten, deren Abfrage und Analyse
 - L => Logstash
(Tool für Datenaufnahme (Data Ingestion), Transformation, etc., wie Flume)
 - K => Kibana (Tool für die webbasierte Visualisierung von Daten)
(nutzbar auch für Qualitätskontrolle der Daten/Geodaten)

GEOJSON

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [125.6, 10.1]
  },
  "properties": {
    "name": "Dinagat Islands"
  }
}
```

Data

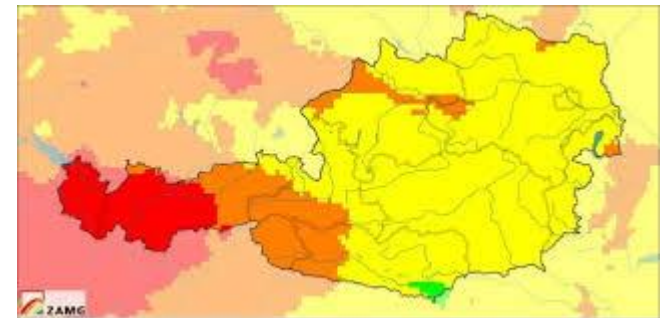


Analyse und Weiterverarbeitung

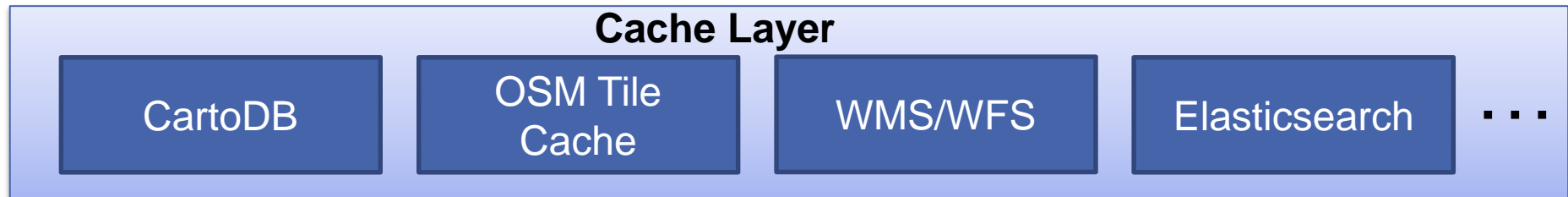
- Service-orientierte Infrastrukturen für die Datenerfassung und das Datenmanagement bieten ideale Andockpunkte für die Analyse und Weiterverarbeitung der Daten
 - Nicht nur für automatisierte Qualitätskontrolle
 - Für die Generierung von abgeleiteten Wissen / Daten für z.B. Entscheidungssysteme
 - Über Aggregation von Daten
 - Bewertung von Daten (Grenzwertüberschreitung)
 - Machine Learning, etc.
- Beispiele
 - Google Traffic Anzeige
 - Ozon- oder Pegelwarnsysteme
 - Waldbrandgefahr
 - Feinstaubalarm
 - Etc.



Google Traffic



Ozonalarmkarte
Österreich



- Endanwendungen benötigen Services, die Daten **schnell und immer verfügbar** ausliefern
- Hierzu sollten Daten in einem Caching-Layer mit geeigneten Services und Datenmanagementsystemen vorgehalten werden
 - Können Cloud-Lösungen (z.B. CartoDB) oder On-Premise-Lösungen sein (z.B. eigener WMS oder Tile Cache)
 - Mix aus einzelnen Services für dedizierte Zwecke
 - Image oder Vektor Tile-Server für Grundkarten (Tile Caches oder WMS)
 - WFS für Feature-Bereitstellung
 - Systeme, wie Elasticsearch für Durchsuchbarkeit, Aggregation and Analytics
- Dedizierte Werkzeuge für Aktualisierung der Cache Inhalte aus „Produktionsdatenmanagement“ notwendig (Cache-Update-Services)

- Sollen Geodaten für unterschiedliche Anwendungen bereitgestellt werden, bietet eine service-orientierte Infrastruktur zur Erfassung, Bearbeitung, Analyse und Bereitstellung der Daten viele Vorteile
 - (Mobile) Erfassung über eine Vielzahl mobiler Anwendungen und Geräte möglich, sofern gängige Formate und APIs instrumentiert werden
 - Agil ausbaubare, modulare, automatisierbare Infrastruktur zur Weiterbearbeitung der Daten
 - Für Bereinigung, Transformation, Anreicherung der Daten
 - Möglichkeiten zur Erzeugung neuen Wissens über Integration von Analyseworkflows
 - Integrierbarkeit von Datenmanagement, Editier- und Kontrollwerkzeugen
 - Caching-Layer erlaubt schnelle Bereitstellung über einen Mix aus externen Services in der Cloud und Diensten im eigenen Rechenzentrum