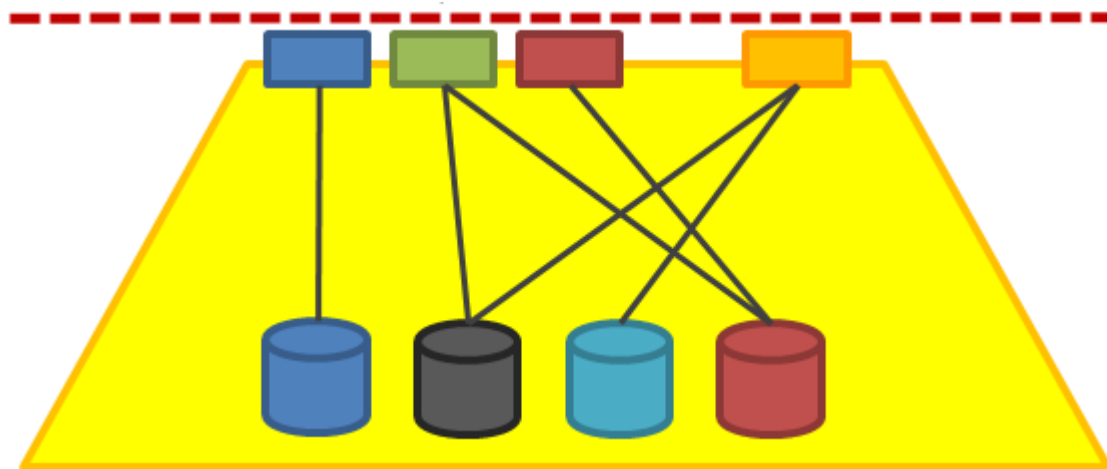
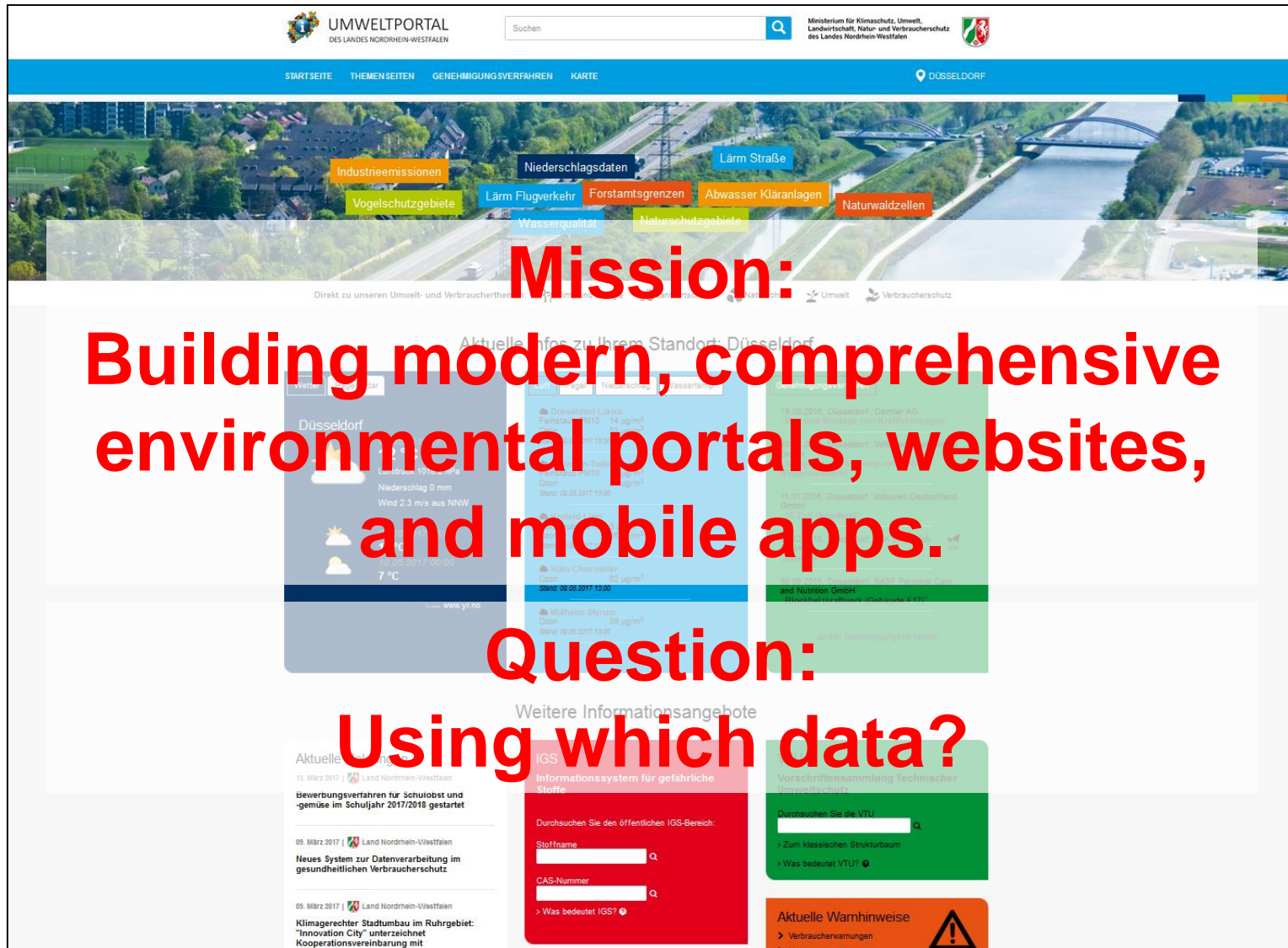


A Generic Web Cache Infrastructure for the Provision of Multifarious Environmental Data

Thorsten Schlachter, Eric Braun, Clemens Döpmeier, Christian Schmitt, Wolfgang Schillinger

INSTITUTE FOR APPLIED COMPUTER SCIENCE (IAI)

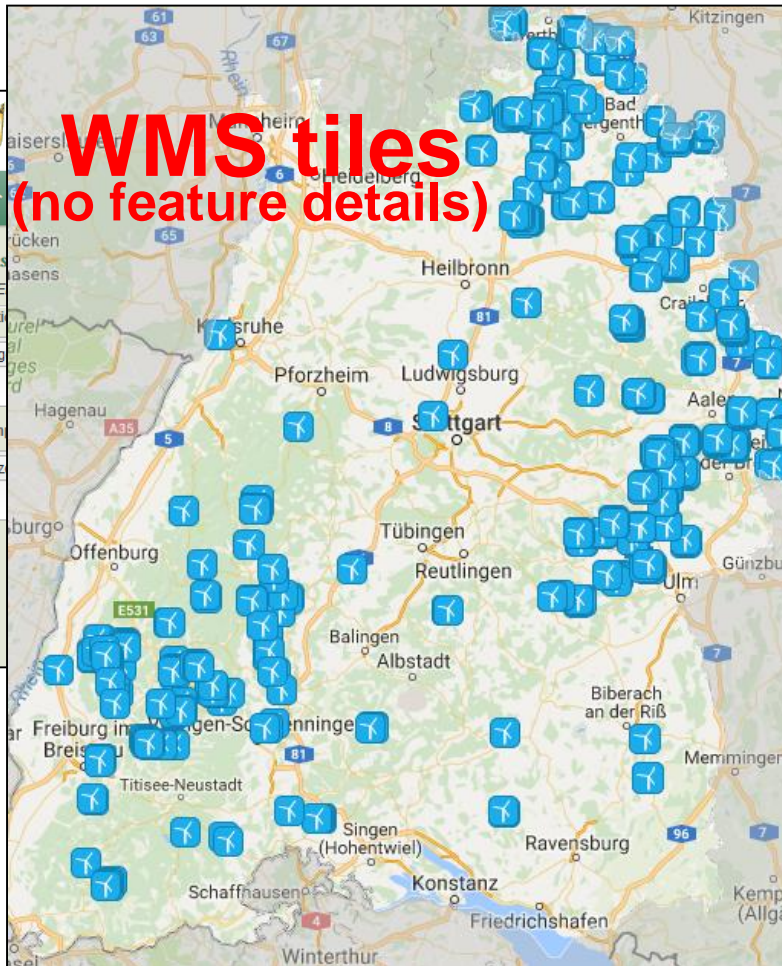




The screenshot shows the 'Umweltportal' website for North Rhine-Westphalia. The header includes the logo, search bar, and navigation links. A blue navigation bar contains 'START SEITE', 'THEMEN SEITEN', 'GENEHMIGUNGSVERFAHREN', and 'KARTE'. The location is set to 'DÜSSELDORF'. The main content area features an aerial view of a city with several data overlays: 'Industrieemissionen', 'Niederschlagsdaten', 'Lärm Straße', 'Vogelschutzgebiete', 'Lärm Flugverkehr', 'Forstamtsgrenzen', 'Abwasser Kläranlagen', 'Wasserqualität', 'Naturschutzgebiete', and 'Naturwaldzellen'. Large red text is overlaid on the page: 'Mission: Building modern, comprehensive environmental portals, websites, and mobile apps.' and 'Question: Using which data?'. The bottom of the page shows a sidebar with news items and a search box for the 'IGS' (Informationssystem für gefährliche Stoffe).

Motivation

- **Example 1: Data available, but not in required representations**



Metadata

Aggregationsart	Aggregationszei...	Berechnungsvorschrift	Datenverfügbarkeit Anfang	Datenverfügbarkeit Ende
Anzahl	1 Jahr	Ozon: 3-Jahresmittelwert ...	01.01.1995 00:00	01.01.2017 00:00
Anzahl	1 Jahr	Ozon: Anzahl der Tage mit ...	01.01.1995 00:00	01.01.2017 00:00
Anzahl	1 Jahr	Ozon: Anzahl der Tage mit ...	01.01.1995 00:00	01.01.2017 00:00
Anzahl	1 Jahr	Ozon: Anzahl der Tage mit ...	01.01.1995 00:00	01.01.2017 00:00

Umweltschutz

zum Ministerium für Umwelt, Klima und Energiewirtschaft

Immissionsdaten > Reg.-Bez. Karlsruhe > Eggenstein

HTML-Tables

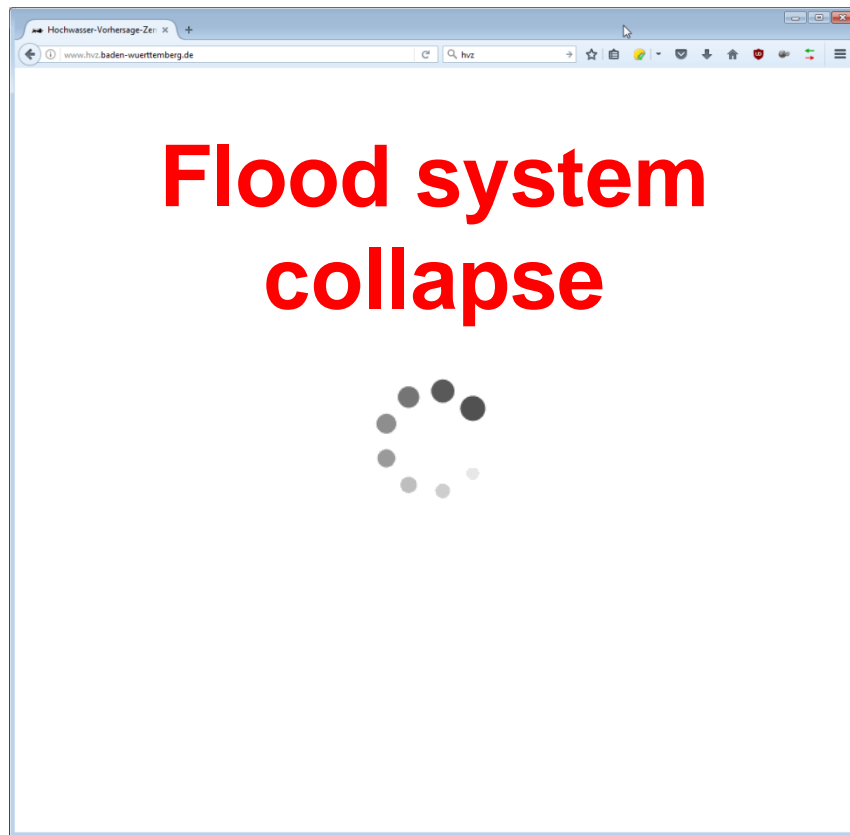
08.05.2017 16:00

	08.05.2017		07.05.2017				
Ortsraum	Aktueller Messwert	Tages-Maximum	Tages-Maximum	Tages-Mittelwert	Immissions-Werte	Anzahl Überschr.	Einheit
Stdt.	25	25	16	10	50 ¹⁾	6	µg/m ³
Stdt.	10	34	19	7	200 ²⁾	0	µg/m ³
Stdt.	51	55	65	46	180 ³⁾		µg/m ³
Stdt.	47	54	51		120 ⁴⁾		µg/m ³
Stdt.	1	3	3	1	350 ⁵⁾	0	µg/m ³

■ Example 2: Services not available during maintenance windows


Motivation

- **Example 3:** Services not scalable; current data not available for depending apps and websites during crisis situations



Motivation

■ Example 4: Legal issues



Login
Anmelden am System

Limited access

Hinweis: Um die Autorenumgebung des Systems verwenden zu können, müssen bei Ihrem Browser *Cookies* und *Javascript* aktiviert sein.

Name:

Passwort:

Falls Sie noch keine Zugangsberechtigung als Autor besitzen, können Sie hier eine entsprechende Benutzerkennung beantragen:

Data protection

```

{
  "contact": "John Doe",
  "position": "CEO"
}
```

(Missing) licences

Datenlizenz Deutschland – Zero – Version 2.0

Jede Nutzung ist ohne Einschränkungen oder Bedingungen zulässig.

Die bereitgestellten Daten und Metadaten dürfen für die kommerzielle und nicht kommerzielle Nutzung insbesondere

1. vervielfältigt, ausgedruckt, präsentiert, verändert, bearbeitet sowie an Dritte übermittelt werden;
2. mit eigenen Daten und Daten Anderer zusammengeführt und zu selbständigen neuen Datensätzen verbunden werden;
3. in interne und externe Geschäftsprozesse, Produkte und Anwendungen in öffentlichen und nicht öffentlichen elektronischen Netzwerken eingebunden werden.

Datenlizenz Deutschland – Namensnennung – Version 2.0

(1) Jede Nutzung ist unter den Bedingungen dieser „Datenlizenz Deutschland – Namensnennung – Version 2.0“ zulässig.

Die bereitgestellten Daten und Metadaten dürfen für die kommerzielle und nicht kommerzielle Nutzung insbesondere

1. vervielfältigt, ausgedruckt, präsentiert, verändert, bearbeitet sowie an Dritte übermittelt werden;
2. mit eigenen Daten und Daten Anderer zusammengeführt und zu selbständigen neuen Datensätzen verbunden werden;
3. in interne und externe Geschäftsprozesse, Produkte und Anwendungen in öffentlichen und nicht öffentlichen elektronischen Netzwerken eingebunden werden.

The Goal

Applications



- A wide variety of applications
 - websites and portals
 - mobile apps
 - professional systems
 - new systems with added values
 - open (government) data, linked data

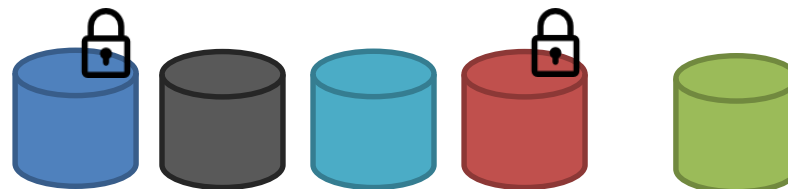
- Using as much data from environmental authorities as possible

The Situation

- Data are available in principle and wait for dissemination
- Limiting factors
 - technical: interfaces, data formats, services
 - professional: semantics, schemas, representations
 - legal: licences, data protection/privacy, liability, using „the Cloud“
 - political: release of data, open (government) data
 - organizational: thinking in projects and systems
 - operational: hosting of systems, versatility
 - economical: efforts, costs
 - cultural: „my data belongs to me“, liability, provenance
 - ...

Data Sources

(complete data, read&write)



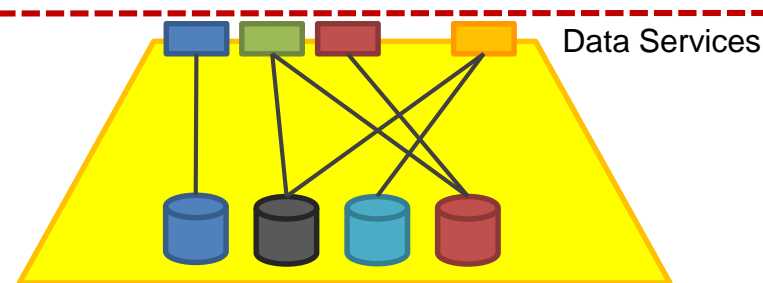
The Idea

- Provide THIS data by means of a service-oriented platform
 - available 24/365
 - scalable upon need (#requests)

SoA

Web Cache

(highly available, scalable)



- clear SLAs as well as usage and exploitation rights
- compliance of data protection and privacy
- documentation of (stable) interfaces, data formats and semantics

Our Solution: The Web Cache Architecture

Applications

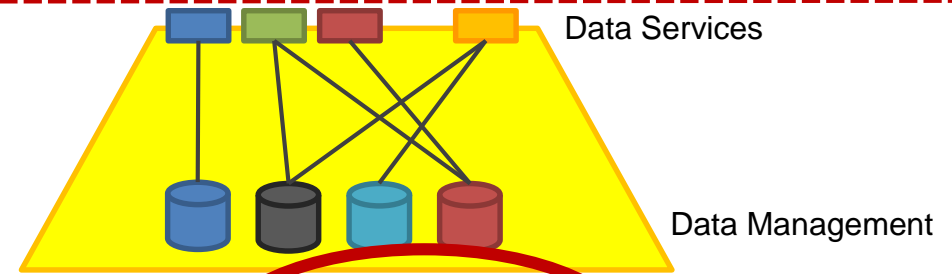
(use of services as needed)



SoA

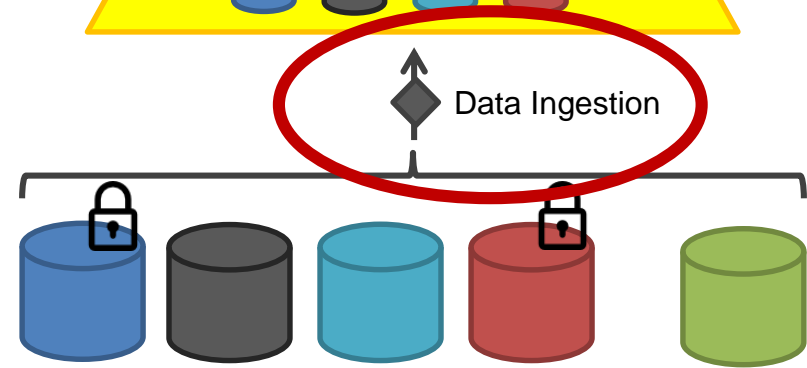
Web Cache

(possibly reduced data, read-only, highly available, scalable)



Data Sources

(complete data, read&write)



The Data Ingestion Process (1)

- Has to tackle most of the "soft" factors
 - professional
 - legal
 - political
 - organizational
 - economical
 - cultural
- Talk to the people with the data
- Talk to decision-makers



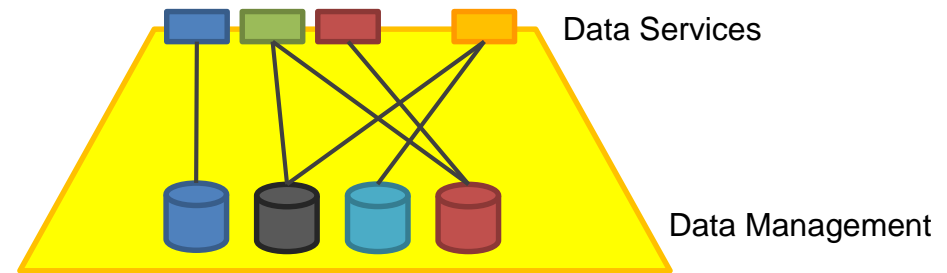
Be patient and insistent!

The Data Ingestion Process (2)

- Technical and operational issues are relatively easy to solve
 - Using suitable tools, e.g.
 - Apache **Flume**
 - Elastic **Logstash**
 - Safe **FME**
 - ...
 - Finding operational providers
 - Cloud providers, e.g. Google, Amazon, etc.
 - Public data centers
 - Administrative data centers

A Stack of Suitable Services

- Collection of generic data services
 - master data
 - schemas
 - time series
 - (media and) digital assets
 - (full text) search
 - spatial data
 - metadata
 - links and relations
- Supplementary services
 - application configuration
 - data discovery



- Suitable persistence systems for data management
 - SQL, e.g. MySQL, Cloud SQL, ...
 - NoSQL, e.g. MongoDB, Elastic, ...
 - Search engines, e.g. Google Search Appliance, Elastic, ...
 - DMS, e.g. Alfresco
 - ...

Düpmeier, C.: **"A Generic Microservice Architecture for Environmental Data Management"** (10:30 Room BURA)

Core Design Principles of the Services

- Easy-to-use **interfaces**: REST APIs
 - Stable URIs, e.g. for use in Semantic Web
 - **Standard** data formats, e.g. RDF, OGC standards, ...
 - **Content negotiation**
- (Common) **schemas** and **vocabularies**
- **Relations** between vastly independent objects
- Implementation as **microservices**
- Use of **container virtualization** technologies
- **Horizontal scalability**, e.g. services, backend systems, platforms
- **Abstraction layers** (adapters) connecting concrete (backend) systems
- **Versioning** of everything, e.g. APIs, data formats, schemas, software modules, data items

Experiences from Running the Web Cache

- Web Cache **in operation** since beginning of 2016
 - gradual start-up
 - straightforward development of single services
 - spatial service still realized with CARTO (www.carto.com)
 - good performance and scalability
 - simple deployment, rolling updates (Kubernetes, Docker)
- **Data ingestion** requires **efforts**
 - many **common patterns**, nevertheless **individual case examination**
 - high level of **automation** possible for operation
- **Significantly more data available** than ever before
 - many new use cases and systems
 - creativity, new ideas in administration

Integrated Search Functionality

halle nsg

Suchen

Orte

Halle (Saale)

Schutzgebiete

Biosphärenreservat

Landschaftsschutzgebiet

Nationalpark

Naturschutzgebiet

RAMSAR Feuchtgebiet

Schutzgebiet (374)

Naturschutzgebiet (346)

Erlebnis (7)

Geoerlebnis (6)

Naturerlebnis (1)

Luftmesswerte

3 Halle/Nord

Feinstaub 10 µg/m³

Ozon 85 µg/m³

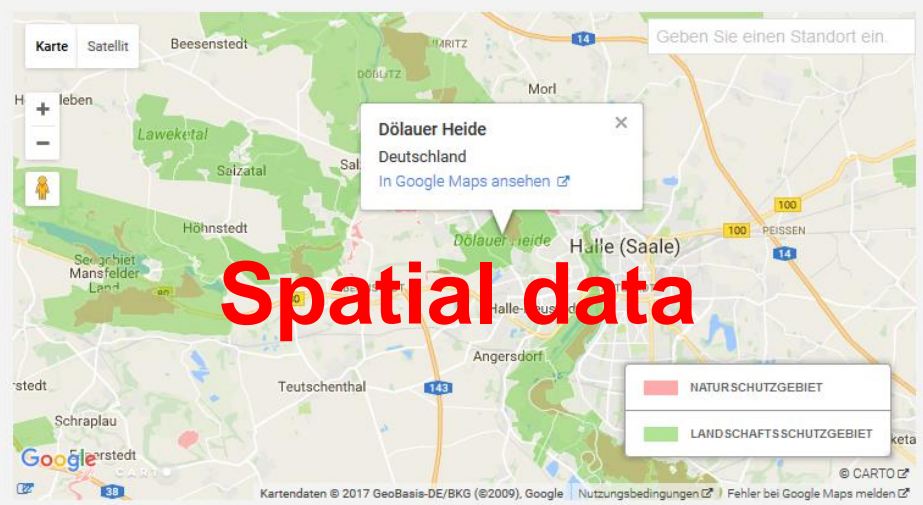
09.05.2017 13:00

Pegelstände

Saale in Halle-Trotha UP

156 cm

09.05.2017 13:00



Spatial data

Landschaftsschutzgebiet

Dölauer Heide

Gebietsnummer: STLSG0037HAL

Suche nach "halle nsg"

Ergebnis 1 - 10 von 777

1 | **Naturschutzgebiete in Sachsen-Anhalt**

https://lvwa.sachsen-anhalt.de/das-lvwa/landwirtschaft-umwelt/naturschutz-landschaftspflege-bildung-fuer-nachhaltige-en-...
 ... A, Abtei und Saaleaue bei Planena, Saalekreis, Stadtkreis: Halle. Acker, ...
 Busch, Harz. ... Bischofwiese, Stadtkreis: Halle. Blonsberg, Saalekreis. ...

2 | **BA 0 -- Geschützte Parks (GP)4 201 -- Land**

https://lau.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/MLU/LAU/Naturschutz/Publikationen/Dateien...
 ... 2 Landesamt für Umweltschutz Sachsen-Anhalt Reideburger Str. 47 06116
 Halle/S. ... nach Landesrecht Naturschutzgebiete (NSG) 198 65.964 3,22 ...

3 | **80 NSG Okertal (Juni 2002)**

https://lau.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/MLU/LAU/Naturschutz/Publikationen/Dateien...

Metadata

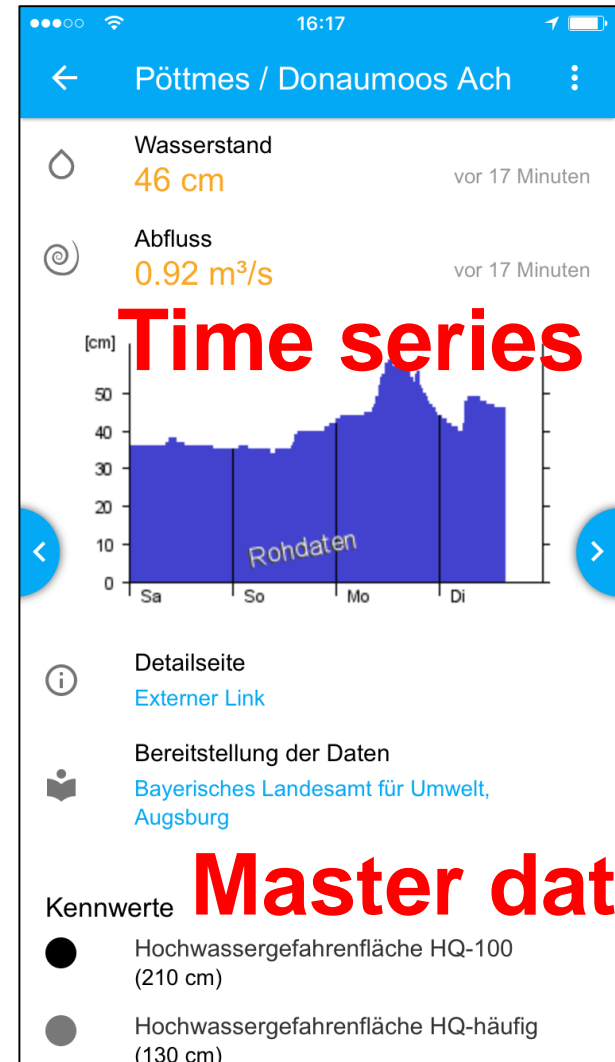
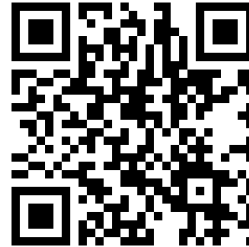
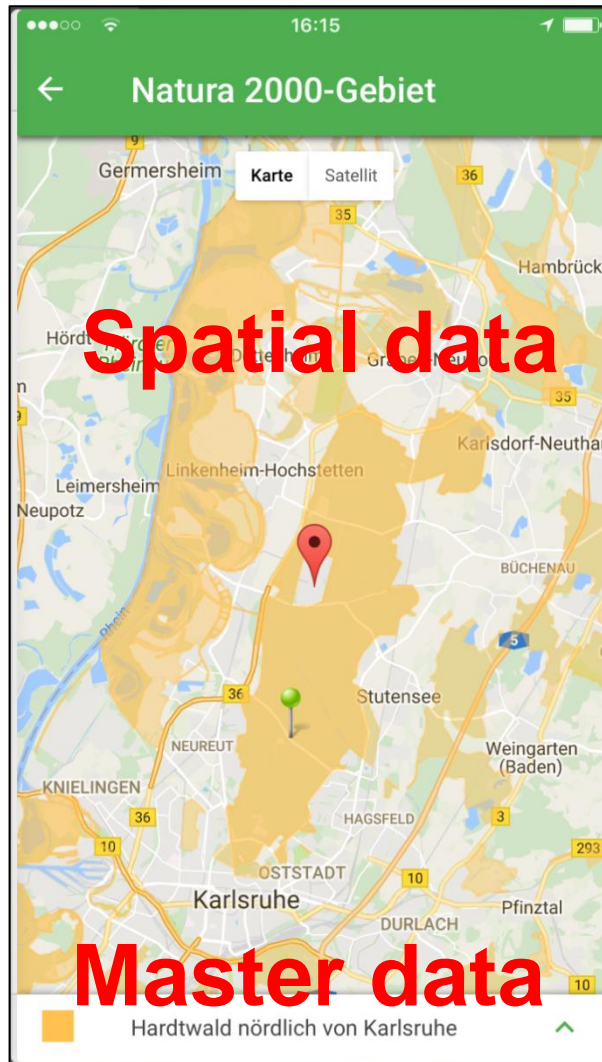
Master data

Time series

Master data

Full text search (assets)

Location-aware Search in Mobile Apps



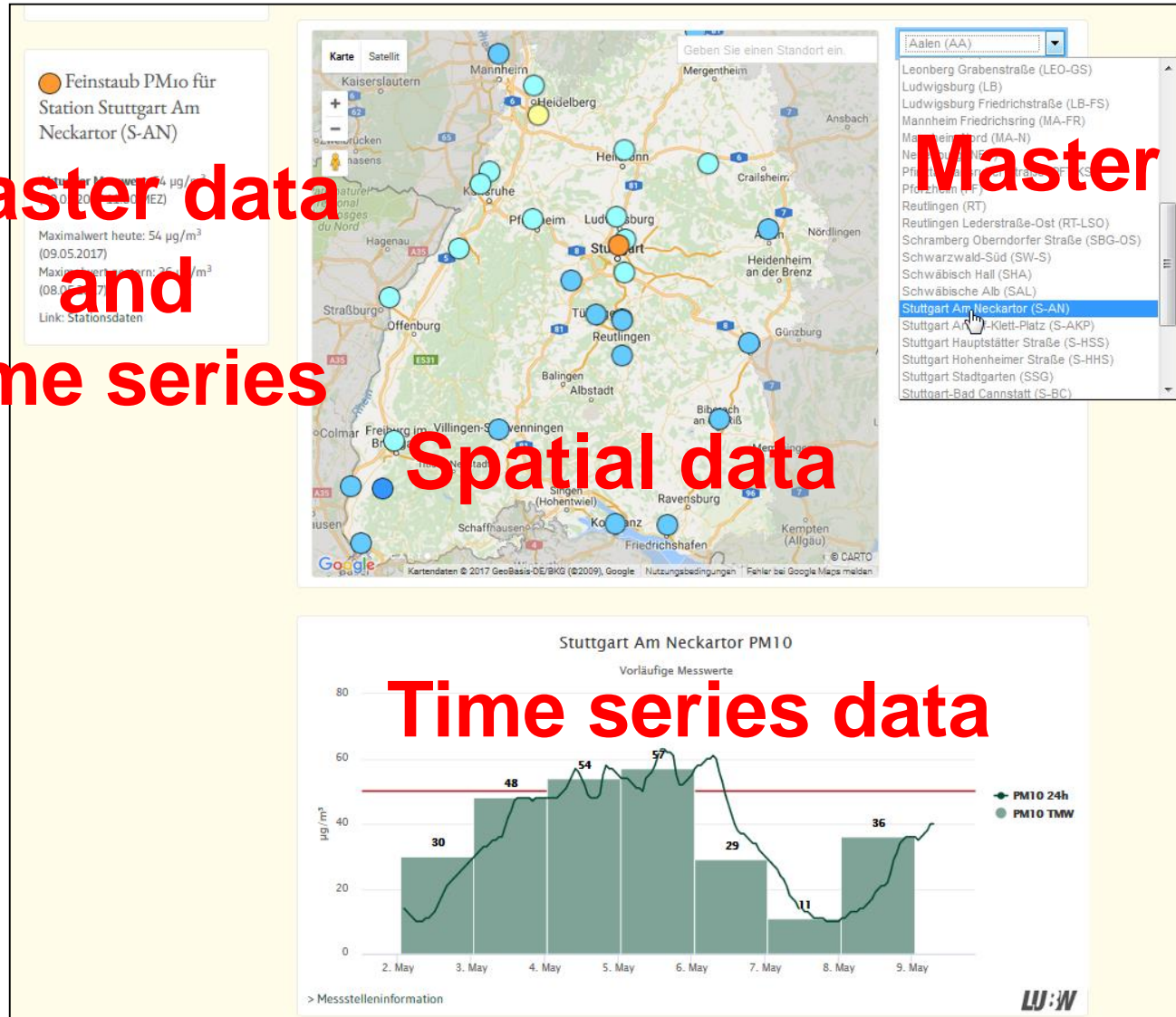
Comprehensive View on Air Quality

**Master data
and
time series**

Master data

Spatial data

Time series data



Conclusion and Outlook

- The Web Cache represents a full architecture for the dissemination of environmental information
 - providing multifarious types of data
 - using a limited set of generic microservices
 - being able to be set up and deployed modularly
 - bringing more environmental information to the people
 - keeping original systems largely untouched
- Development of a set of corresponding generic Web components (HTML5) for use in frontend systems
- Evaluation of use cases having bidirectional data flow, e.g. including data from crowd sourcing apps
- Applications with harder real-time requirements (energy systems)

Thank you for your attention!

contact:

Thorsten Schlachter
Karlsruhe Institute of Technology
thorsten.schlachter@kit.edu
+49-721-608-25769



Credits

■ Images and Icons

- http://www.flaticon.com/free-icon/lock-icon_26053
- <https://pixabay.com/de/besprechung-meeting-gespr%C3%A4ch-1020166/>