

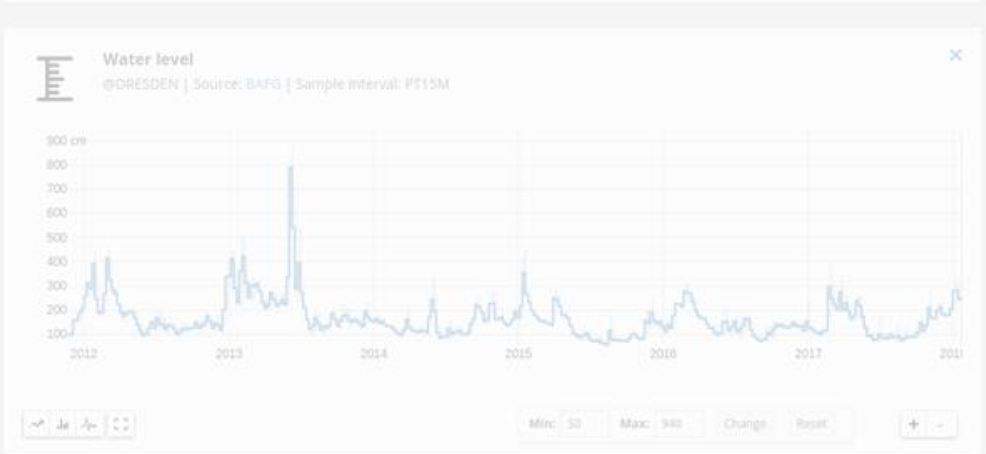
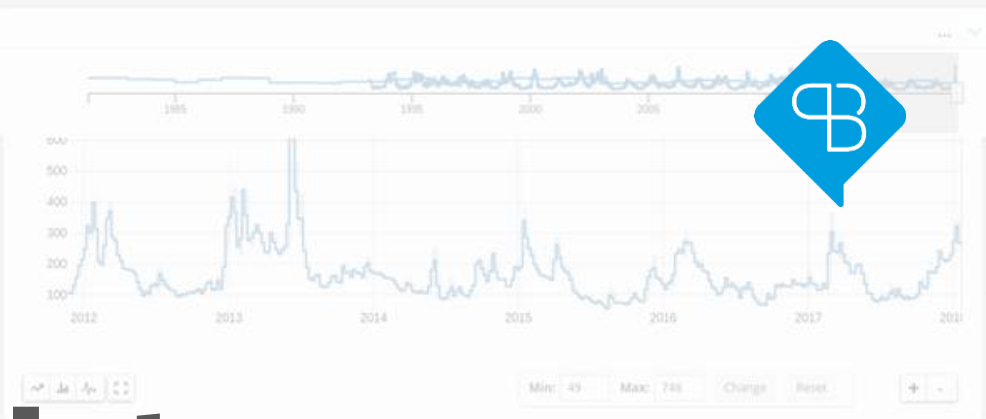
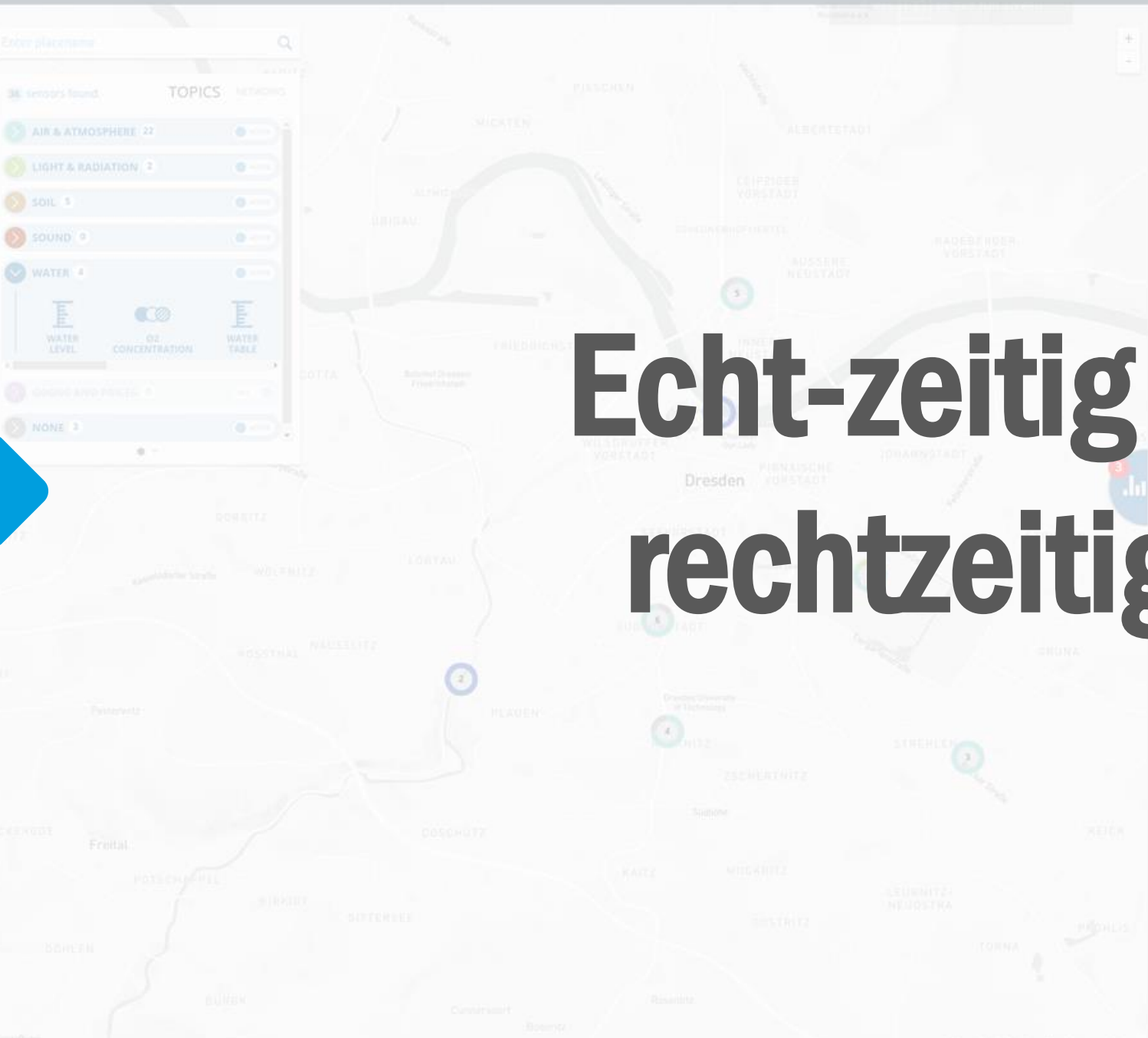


PIKobytes
DATA-DRIVEN SOFTWARE AND CONSULTING.

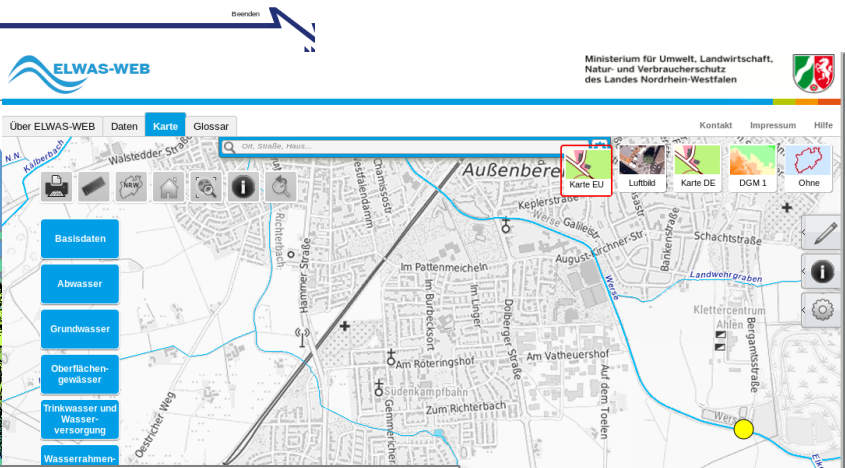
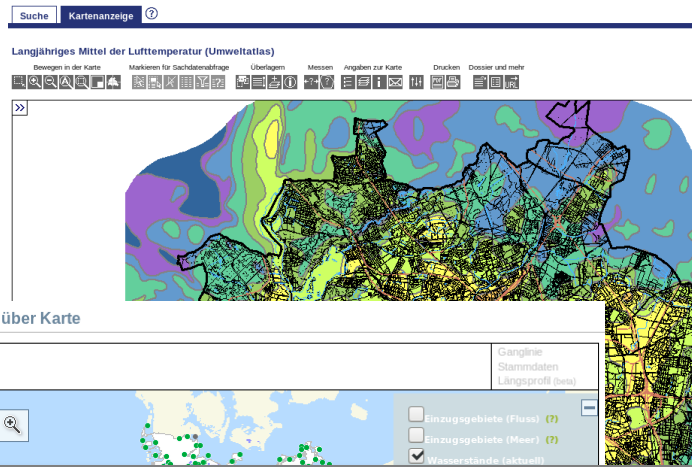
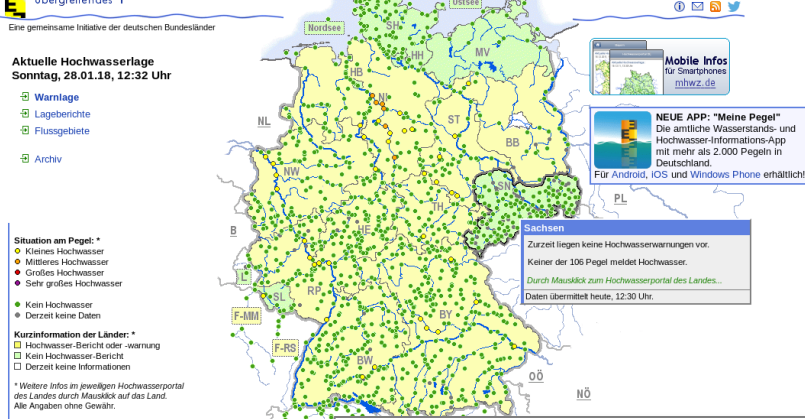
OpenSensorWeb – Effizienter Zugriff auf lokale Umweltsensordaten

Matthias Müller





Echt-zeitig ist rechtzeitig!



“The ultimate goal with Big Data is to collect, store, find meaning, and extract value from data. [...] As things stand now, the data ecosystem is highly fragmented. Between those who create data and those who could potentially extract value from it sits a labyrinth fraught with complexity, disparity, and miscommunication.” – Cisco IBSG, 2012

www.hochwasserzentralen.r

23 Datensätze

Sortieren nach Letzte Änderung

Filter

Suche nach Datensätzen ...

Ansicht

- Analysen 14
- Karte 10
- Benutzerdefinierte Ansicht 1

Letzte Änderung

2018 23

Herausgeber

LHP 22

Ville de Versailles 1

Nutzungsbedingungen der LHP

Kindertagesstätte Kinderbetreuung Potsdam Kita

Grundschulen in Potsdam

Potsdams Grundschulen mit Adressen und Geopunkten

Herausgeber LHP

Lizenz Nutzungsbedingungen der LHP

Grundschule Schule Bildungseinrichtung Potsdam

Grundschulen und ihre Einzugsgebiete

Stadtteile in Potsdam

Potsdams Stadtteile als Geoflächen

Herausgeber LHP

Tagesmitteltemperatur für den La Oberhavel vom DWD

Aktualität der Metadaten: 23.03.2015 00:00

Tagesmitteltemperatur für den Landkreis i vom DWD. Die Tagesmitteltemperatur, au Tagesdurchschnittstemperatur, wird seit d 2001 durch Bilden des Mittelwerts aller zu Stunde gemessenen Lufttemperaturwerte

2 Detailsicht XML

Tagesmitteltemperatur für den La Oberhavel vom DWD als Rasterdati

Aktualität der Metadaten: 29.06.2015 00:00

Tagesmitteltemperatur für den Landkreis i vom DWD als Rasterdatensatz. Die Tagesmitteltemperatur, auch Tagesdurchschnittstemperatur, wird seit d 2001 durch Bilden des Mittelwerts aller zu

3 Detailsicht XML

Wassergütemessnetz 2 des Land Brandenburg

Aktualität der Metadaten: 01.03.2017 00:00

Das Wassergütemessnetz 2 (WGMN2) ist Rahmen der nationalen und international Mediepflichten aktuelle Daten der interess Öffentlichkeit zur Verfügung. Bürger, Schu Behörden haben ein reges Interesse ...

- COOPDaily_announcement_042011.pdf
- COOPDaily_announcement_042011.rtf
- figures
- ghcnd_all.tar.gz
- ghcnd-countries.txt
- ghcnd_gsn.tar.gz
- ghcnd_hcn.tar.gz
- ghcnd-inventory.txt
- ghcnd-states.txt
- ghcnd-stations.txt
- ghcnd-version.txt
- grid
- gsn
- hcn
- isd
- nws
- papers
- readme.txt
- stage1
- stage2
- status.txt
- superghcnd

Größe	Zuletzt verändert
	27.01.2018 23:56:00
	02.01.2018 22:29:00
34 KB	20.04.2011 01:00:00
123 KB	20.04.2011 01:00:00
67 KB	20.04.2011 01:00:00
	06.02.2013 00:00:00
3034573 KB	27.01.2018 20:58:00
4 KB	23.06.2015 01:00:00
147593 KB	27.01.2018 20:58:00
286738 KB	27.01.2018 20:58:00
27080 KB	20.01.2018 20:31:00
2 KB	16.05.2011 01:00:00
8852 KB	24.01.2018 13:37:00
1 KB	27.01.2018 20:58:00
	27.01.2018 20:03:00
	27.01.2018 20:59:00
	27.01.2018 21:01:00
	21.03.2017 00:00:00
	20.10.2017 16:03:00
	02.10.2012 01:00:00
24 KB	11.09.2015 01:00:00
	18.05.2017 01:00:00
	11.05.2017 01:00:00
31 KB	22.01.2018 09:46:00
	28.01.2018 01:57:00

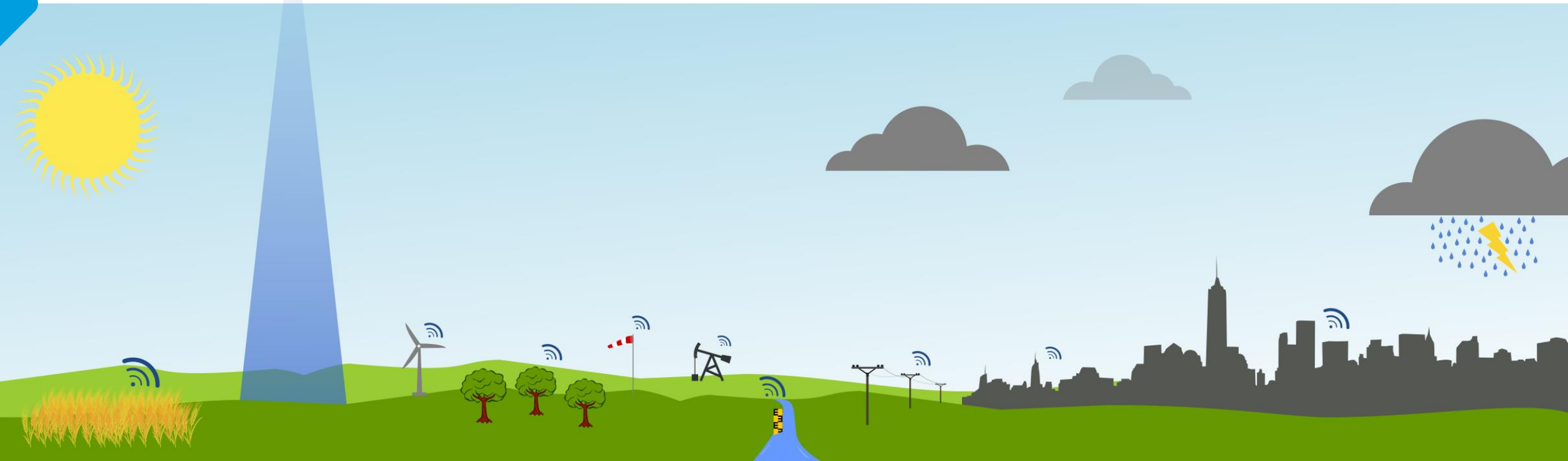
Index of /

https://opendata.dwd.de

Index of /

../		
climate/	06-Jul-2017	10:00
climate_environment/	11-Dec-2017	20:46
test/	04-Jan-2018	10:20
weather/	28-Jan-2018	12:10
README.txt	25-Jul-2017	07:00

Digitale Abbilder der (Um)Welt





OpenData – status quo

Anbieterorientiert statt nachfrageorientiert

- Viele einzelne Portale, viele Dienste, viele Meta-Kataloge
- Räumliche, föderale, institutionelle und **inhaltliche Überlappung**
- Daten müssen nach wie vor **manuell** integriert werden
 - **Kataloge statt Daten**
 - Uneinheitliche Beschreibung / Systematiken / Metadaten
 - Skalierungsprobleme bei Dienste und Schnittstellen für nachgelagerte Applikationen
 - Freiheitsgrade bei Datenformaten und Schnittstellen verhindern vollautomatische Integration
 - **Uneinheitliche Nutzungsrechte und Lizenzen**



Worum geht es?

Messdaten / In-situ-Sensordaten

1. Anbieterübergreifendes Datenangebot
2. **Performantes Datenhandling für Zeitreihen / Messdaten**
3. Angemessene Betriebskosten für ein *heißes* Datenarchiv



<https://commons.wikimedia.org/wiki/File:Luftmesstation.jpg>

Datenanalyse



Wo liegen die nächsten Wetterstationen (Niederschlag, Temperatur, Globalstrahlung) mit Messwerten in stündlicher Auflösungen für den 23.03.2017?

> Multidimensionale Recherche (Raum, Zeit, Thematik, Auflösung, ...)

Wie sehen langjährige Temperatur- und Niederschlagsverläufe am Standort aus? Gibt es zeitliche und räumliche Muster? Was sind Mittel- und Extremwerte?

> Längsschnitte (Charts) und Querschnitte (räumliche Interpolation)



Datenanalyse vs Datenmanagement

Wo liegen die nächsten Wetterstationen (Niederschlag, Temperatur, Globalstrahlung) mit Messwerten in stündlicher Auflösungen für den 23.03.2017?

> Multidimensionale Recherche (Raum, Zeit, Thematik, Auflösung, ...)

Wie sehen langjährige Temperatur- und Niederschlagsverläufe am Standort aus? Gibt es zeitliche und räumliche Muster? Was sind Mittel- und Extremwerte?

> Längsschnitte (Charts) und Querschnitte (räumliche Interpolation)

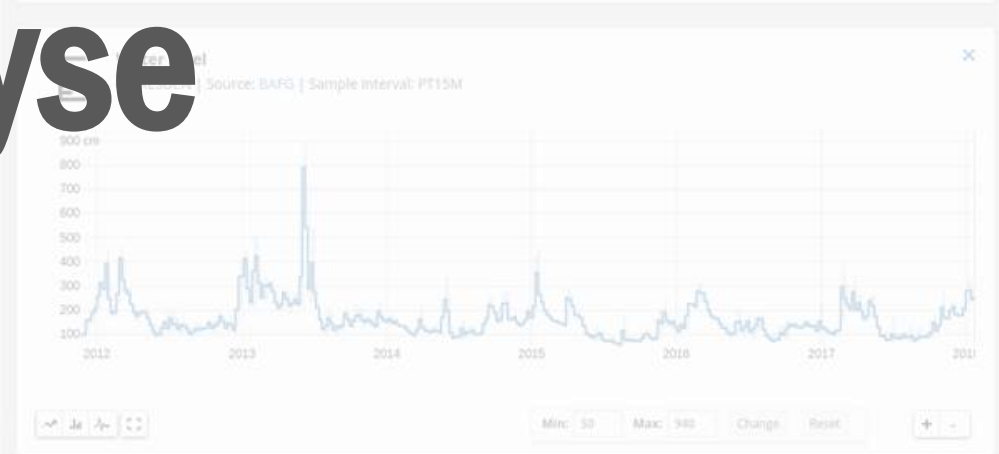
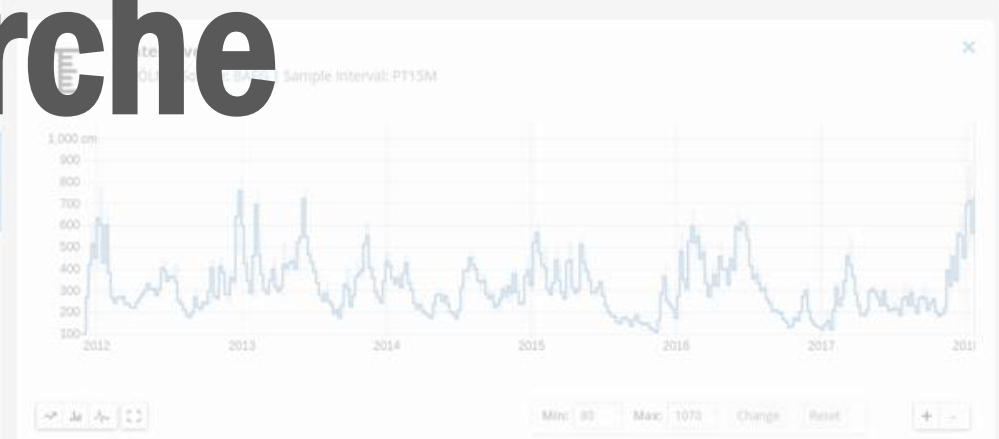
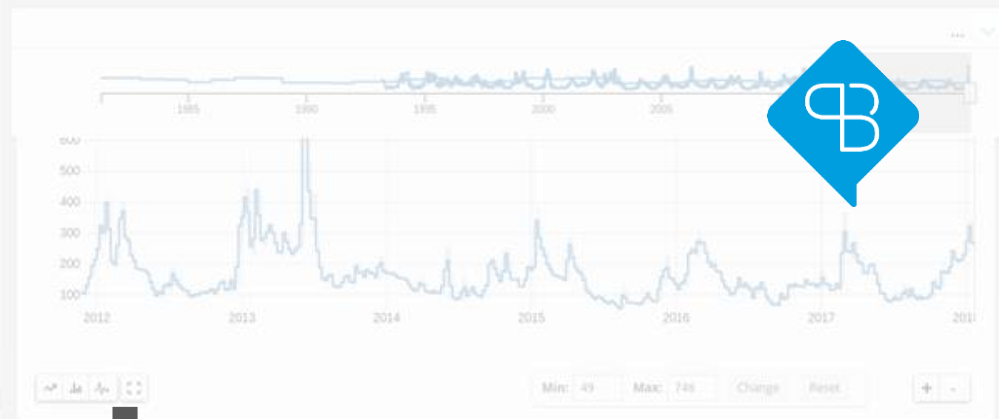
Wie lassen sich viele Daten effizient schreiben, aktualisieren und für den Livezugriff speichern?

> Sensoren im Feld, Synchronisation mit externen Datenarchiven

> Hardwarekosten



Datenrecherche und Datenanalyse

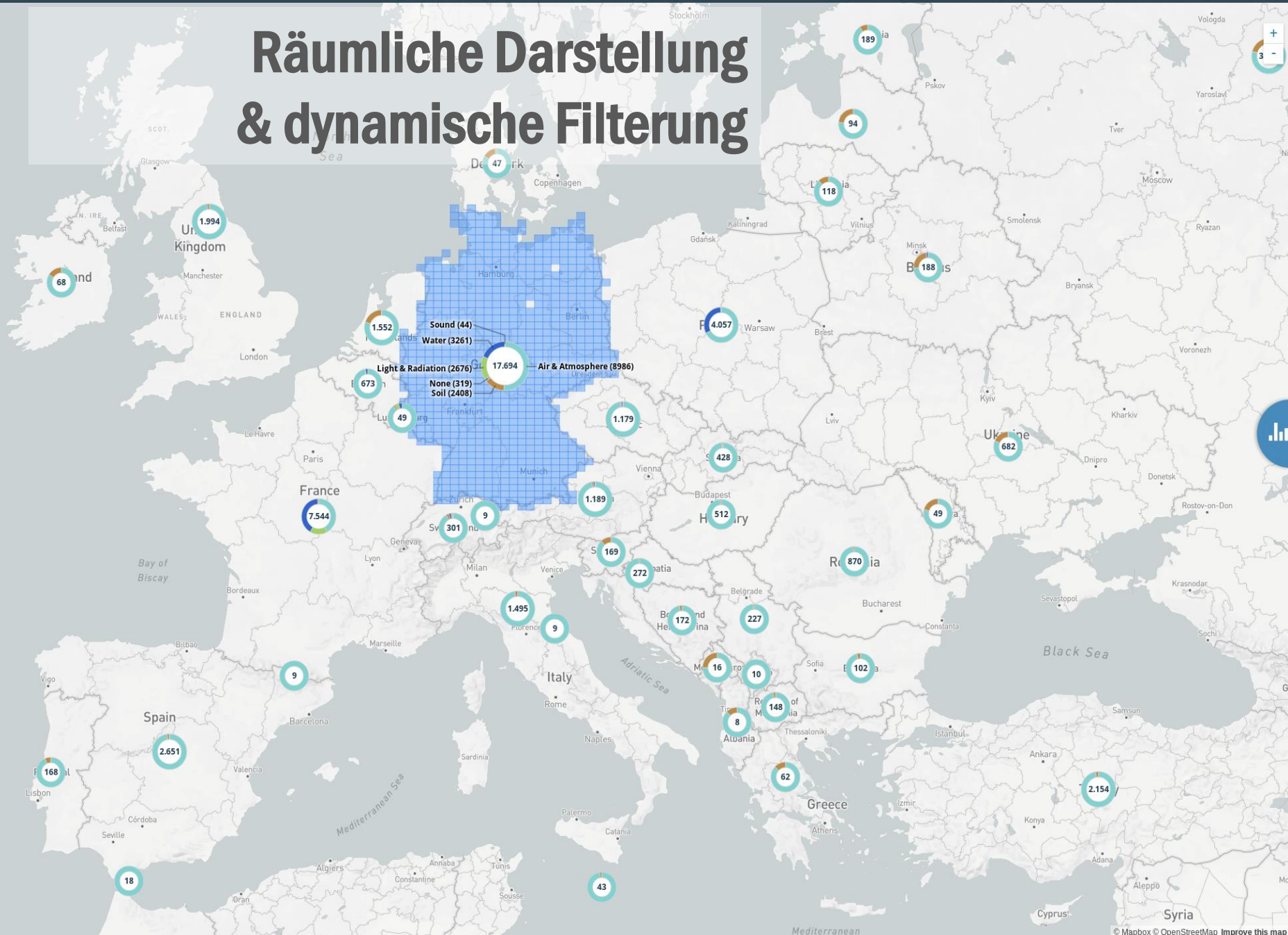


Enter placename

51.180 sensors found. TOPICS NETWORKS

- AIR & ATMOSPHERE 34.962 ACTIVE
- LIGHT & RADIATION 3.929 ACTIVE
- SOIL 4.182 ACTIVE
- SOUND 44 ACTIVE
- WATER 7.741 ACTIVE
- GOODS-AND-PRICES 0 OFF
- NONE 322 ACTIVE

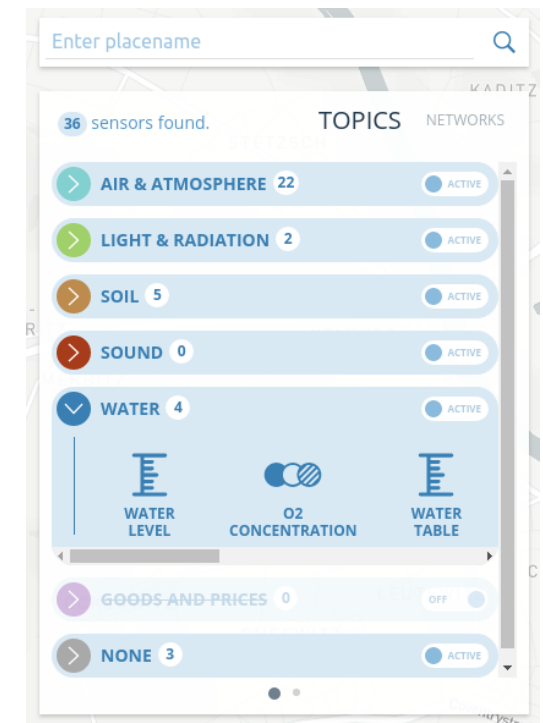
Räumliche Darstellung & dynamische Filterung





Geografische und thematische Filterung

- **Multidimensionale Sensorinformationen**
 - Geo-Koordinaten, Zeitraum, zeitl. Auflösung, Phänomen, Anbieter ...
 - 400.000 Sensoren, 0.5-1.0kB Metadaten: $\geq 200\text{MB}$
- **Clientseitige Filterung unmöglich**
- **Dynamische Datenbankabfrage: ca. 2-8 Sekunden**
- **In-Memory-Indizes (ES, Lucene): ca. 50-200 ms**



Standardvokabulare für Phänomene



Thesauri (GEMET, USGS, UBA ...)

- Nie vollständig
- Selten mehrsprachig
- Unklare Perspektive
 - Aktiv oder tot?
 - Mitwirkung / Ergänzung?
 - Langfristige Verfügbarkeit?

EIONET
European Environment Information and Observation Network

SERVICES | REPORTNET | TOOLS
TOPICS (ETCS)

You are here: Eionet » GEMET English

GEMET General Multilingual Environmental Thesaurus

wind

Found 6 results for 'wind' in English:

- wind
broader context: meteorological phenomenon
- wind erosion
broader context: erosion
- wind power
broader context: non-conventional energy
- wind power station
broader context: electric power plant

EIONET
European Environment Information and Observation Network

SERVICES | REPORTNET | TOOLS
TOPICS (ETCS)

You are here: Eionet » GEMET English

GEMET General Multilingual Environmental Thesaurus

wind direction

Found 0 results for 'wind direction' in English:

Pages: 1

GEMET | DOWNLOAD | ALPHABETS | ABOUT GEMET | WEB SERVICES | DEFINITION SOURCES | HISTORY OF CHANGES
GEMET - Concepts, version 4.0.1, 28 Jun 2017

Document last modified: 2017/08/29 . Content in this portal is modified daily by a community of providers.
Disclaimer. Copyright:

EIONET
European Environment Information and Observation Network

SERVICES | REPORTNET | TOOLS
TOPICS (ETCS)

You are here: Eionet » GEMET English

GEMET General Multilingual Environmental Thesaurus

direction

Found 0 results for 'direction' in English:

Pages: 1

GEMET | DOWNLOAD | ALPHABETS | ABOUT GEMET | WEB SERVICES | DEFINITION SOURCES | HISTORY OF CHANGES
GEMET - Concepts, version 4.0.1, 28 Jun 2017

Document last modified: 2017/08/29 . Content in this portal is modified daily by a community of providers.
Disclaimer. Copyright:

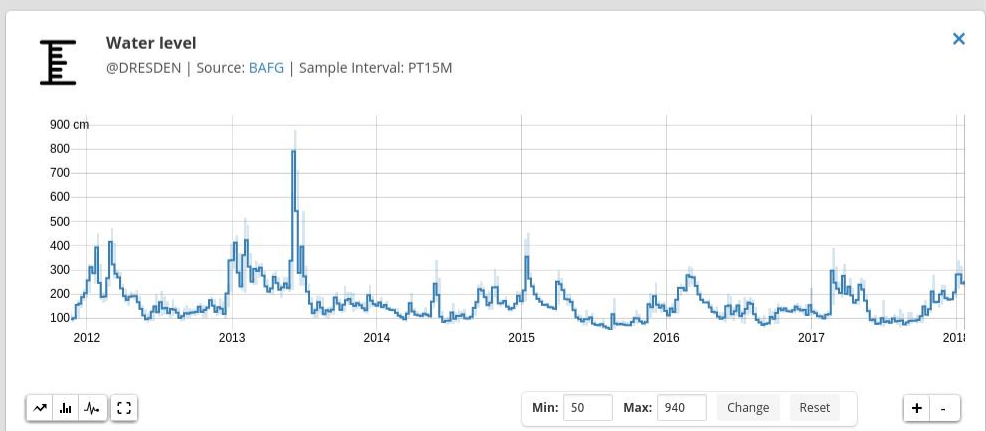
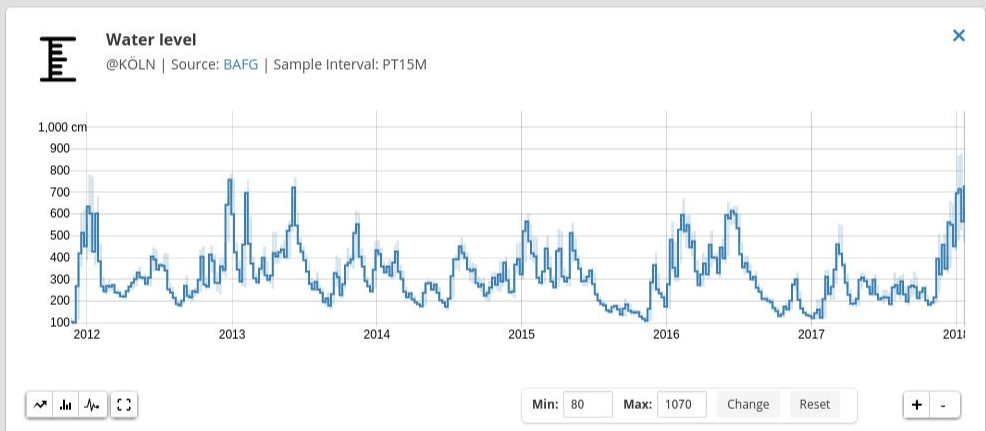
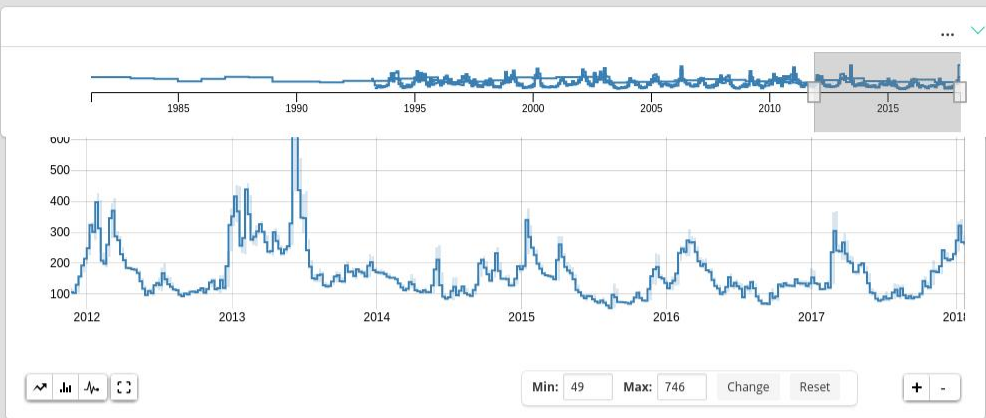
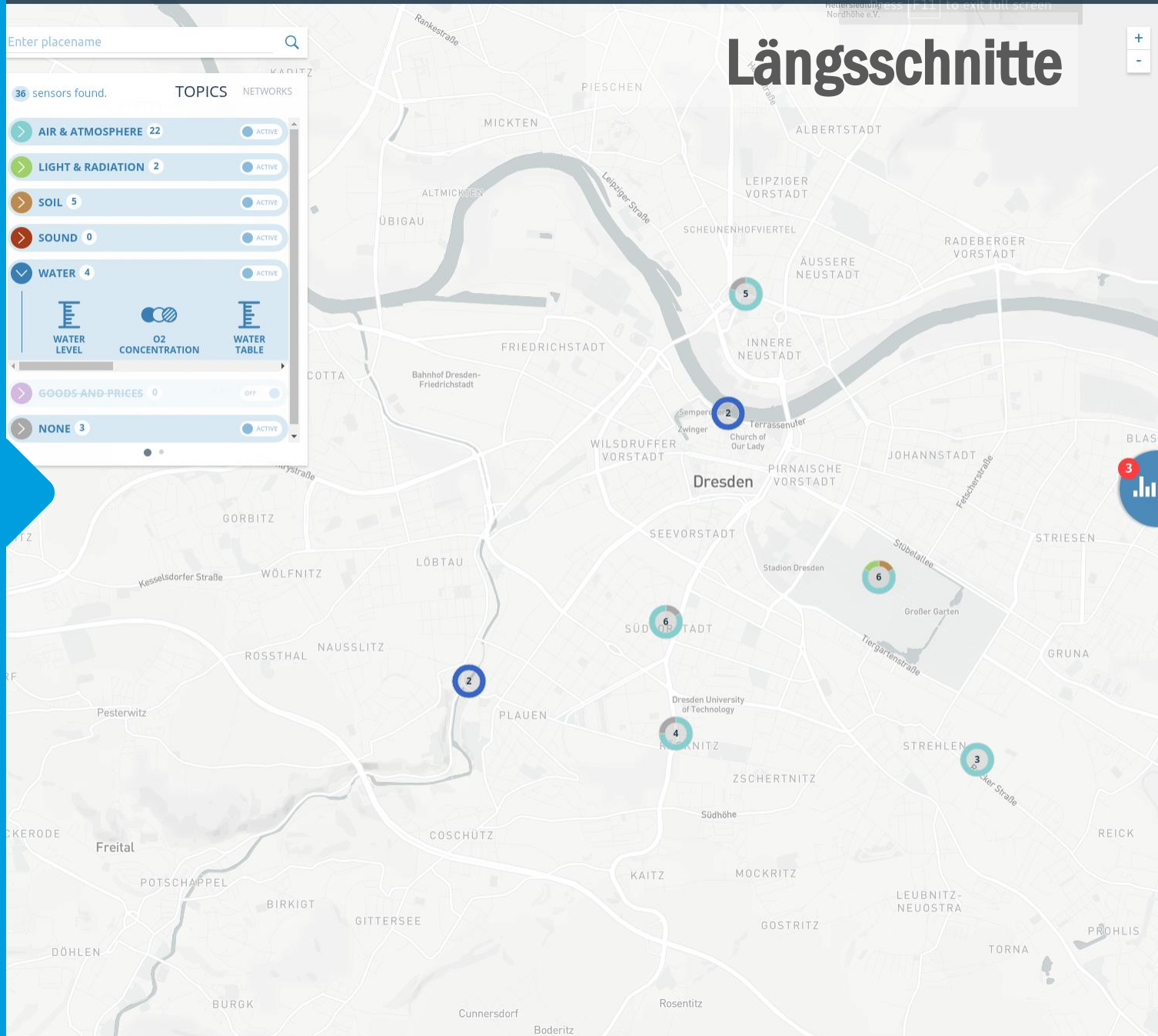
Enter placename

36 sensors found.

TOPICS NETWORKS

- AIR & ATMOSPHERE** 22 ACTIVE
- LIGHT & RADIATION** 2 ACTIVE
- SOIL** 5 ACTIVE
- SOUND** 0 ACTIVE
- WATER** 4 ACTIVE
 - WATER LEVEL
 - O2 CONCENTRATION
 - WATER TABLE
- GOODS AND PRICES** 0 OFF
- NONE** 3 ACTIVE

Längsschnitte





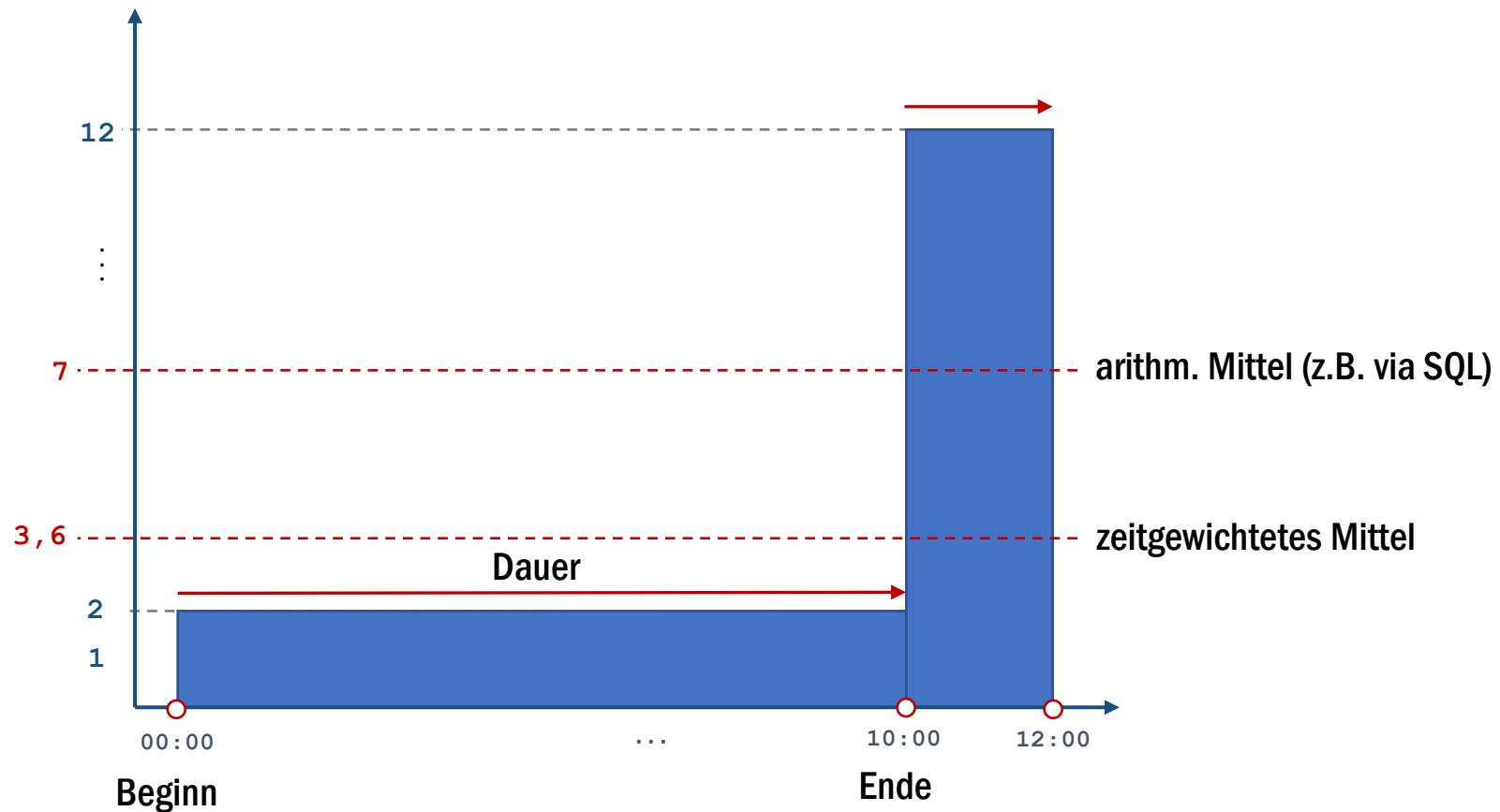
Sequentielles Lesen

- **Streaming von Rohdaten (beliebiger Größe)**
 - Typisch: 100k – 1M Datenpunkte pro Sensor
 - Extrem: 10M – 100M Datenpunkte und mehr
- **Resampling / Regularisieren**
 - Für Diagramme, Vergleichsansichten und Analysen
 - Mittelwert / Min / Max sowie Wertdichte
 - Dynamisches Konvertieren von Zeitzonen
 - Häufigkeit und Dauer der Einzelmessung wichtig
 - Exakte Vorgeneralisierung schwierig



Zeitreihenarithmetik (1)

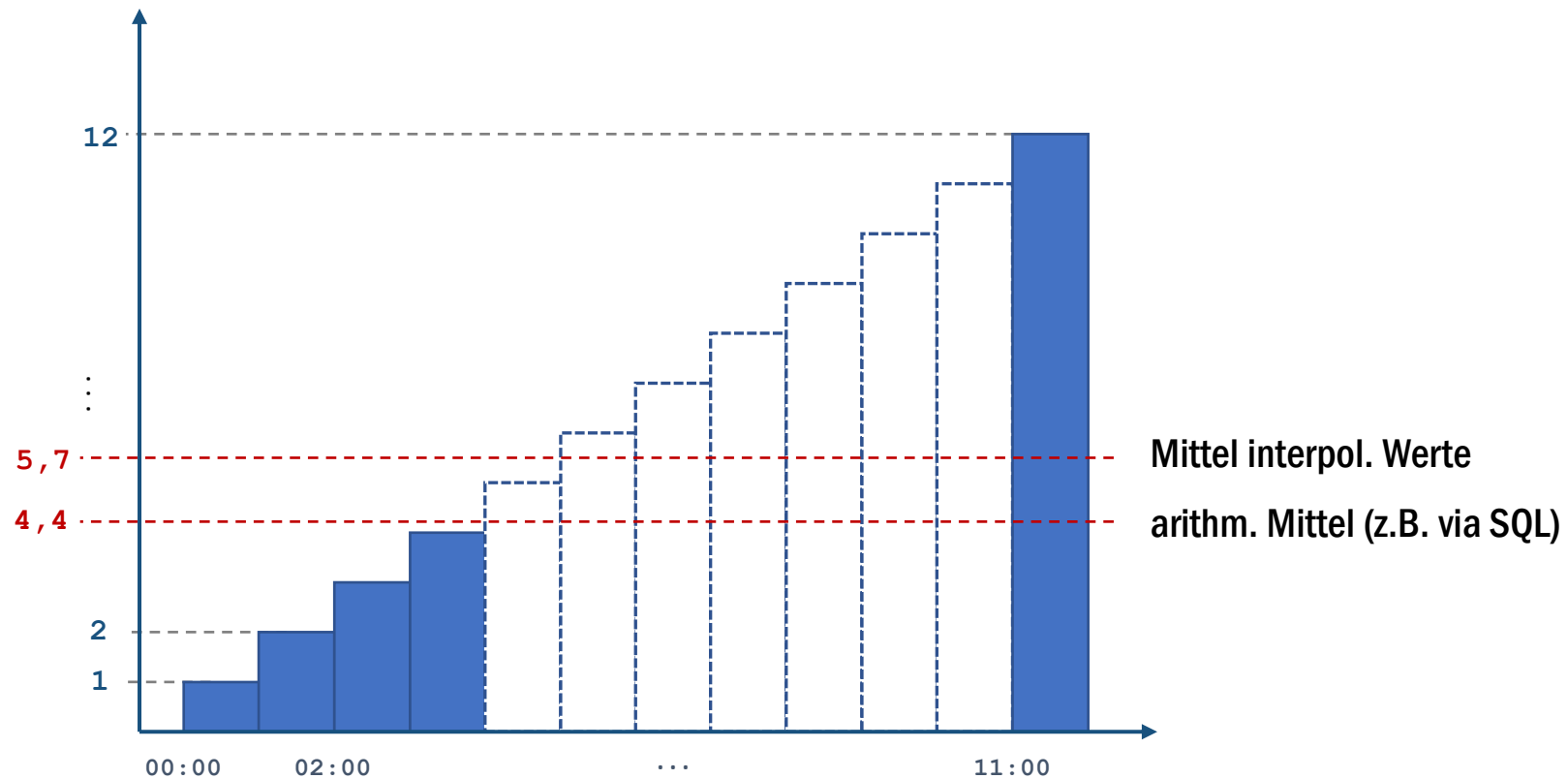
Mittelwertbestimmung für ungleiche Messintervalle





Zeitreihenarithmetik (2)

Mittelwertbestimmung bei Datenlücken



Arctic Ocean
Enter placename

567 sensors found.

TOPICS NETWORKS

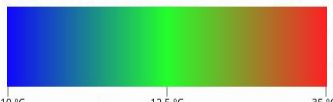
- AIR & ATMOSPHERE 350 ACTIVE
- LIGHT & RADIATION 30 ACTIVE
- SOIL 64 ACTIVE
- SOUND 0 ACTIVE
- WATER 118 ACTIVE
- GOODS AND PRICES 0 OFF
- NONE 5 ACTIVE

Data product:

Air temperature (°C)

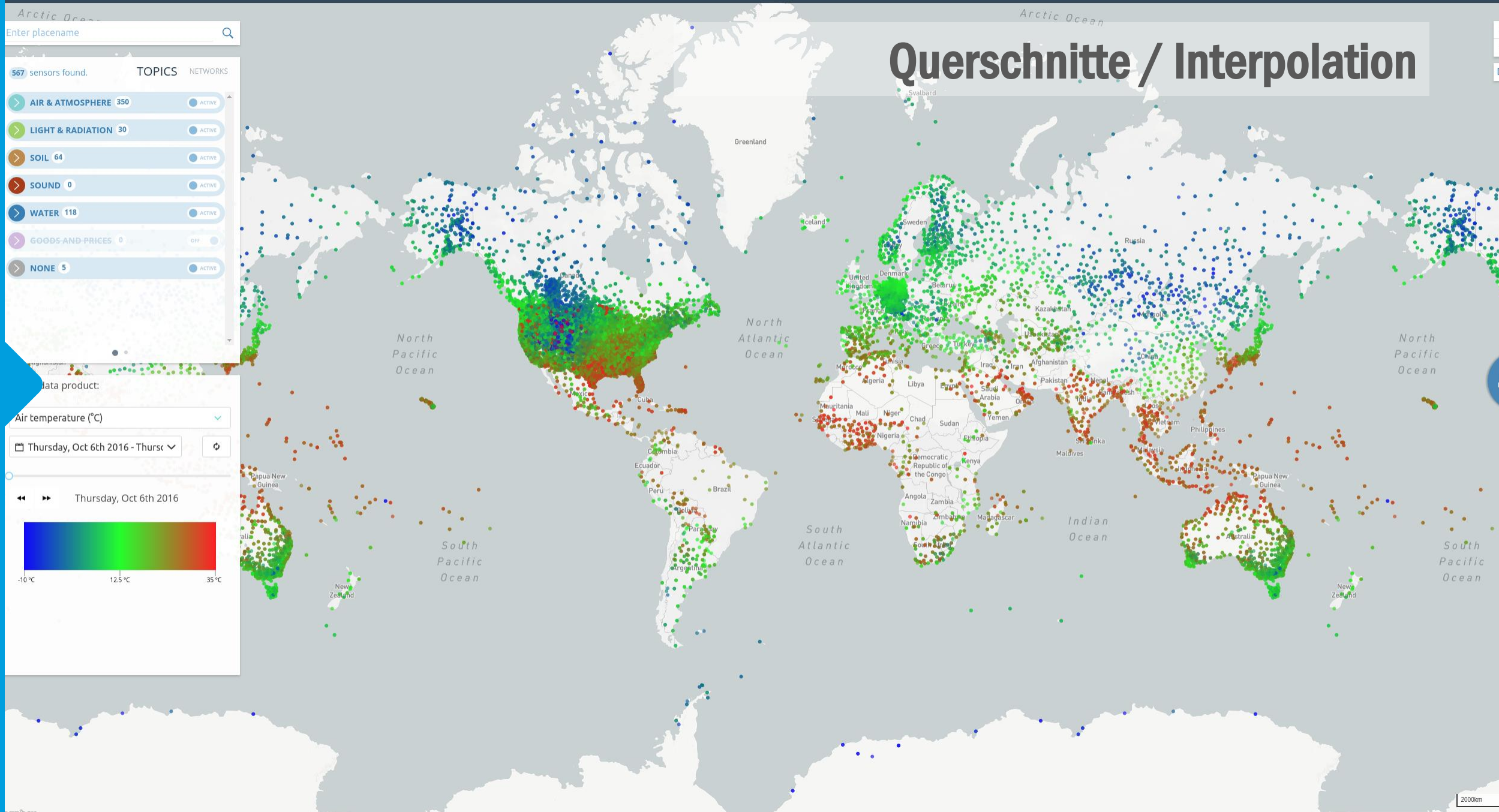
Thursday, Oct 6th 2016 - Thurs

Thursday, Oct 6th 2016



-10 °C 12.5 °C 35 °C

Querschnitte / Interpolation





Räumliche Querschnitte

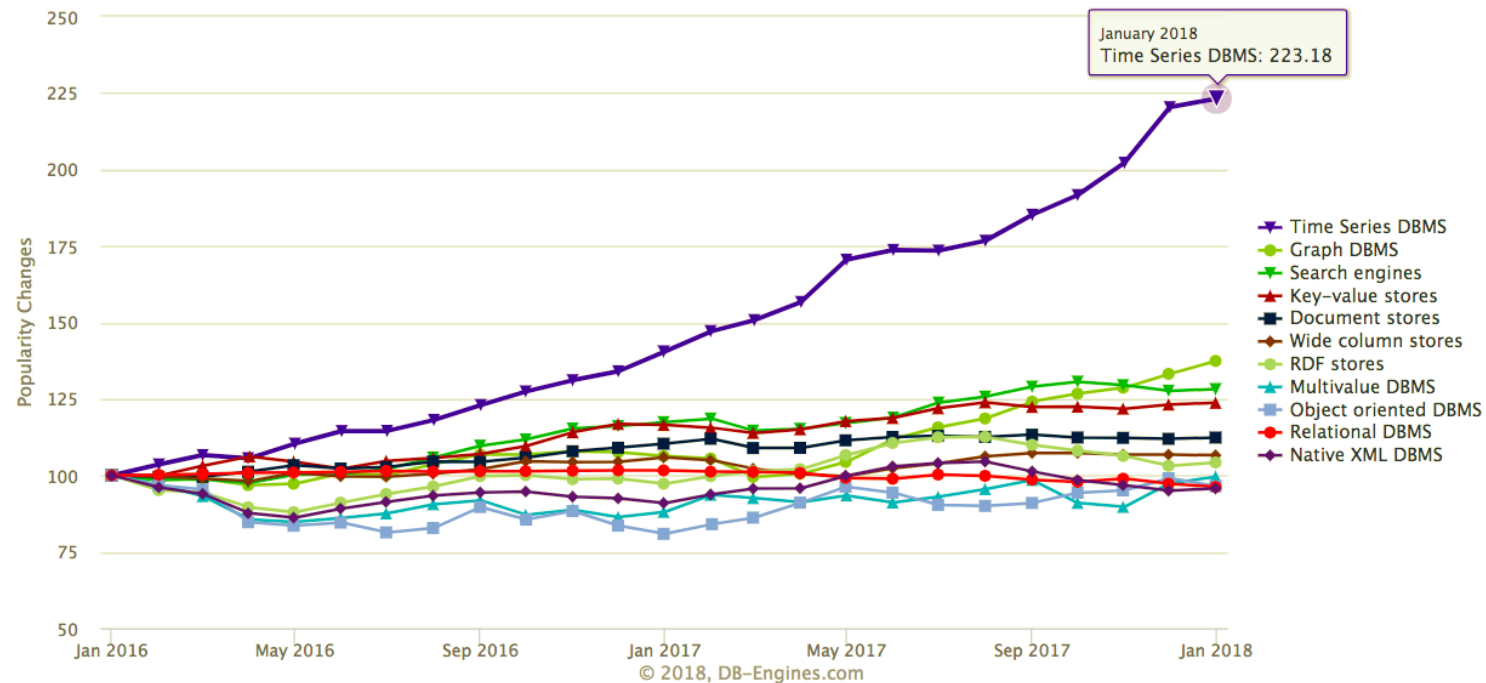
- **Beispiel Temperaturinterpolation**
 - 50.000 Sensoren
 - Bildung aller Tagesmittelwerte für einen Monat (ein Jahr, ...)
 - Vergleichsweise rechenintensiv!

- **Unterschiedliche Messauflösung der Sensoren**
 - Dynamische Regularisierung nötig (Zeitzone, Auflösung)
 - SQL Joins reichen nicht aus

Time Series Databases



Trend of the last 24 months

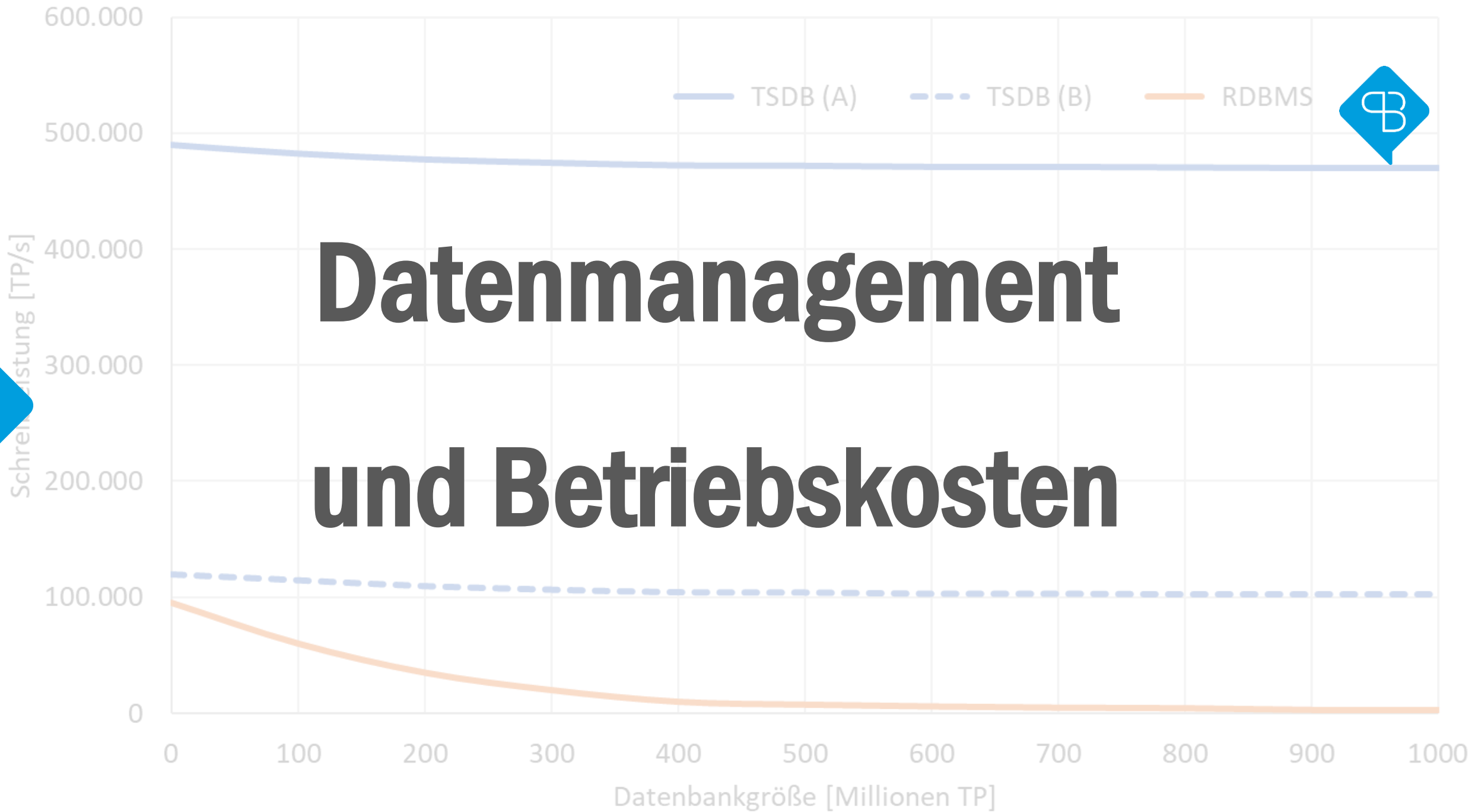


https://db-engines.com/en/ranking_categories

TSDB-Features

- (+) Auf Zeitreihen optimiert
- (+) Zeitreihenstatistik, AKF, ...
- (+) Schnelles append
- (o) Gelegentlich mit Geo-Support
- (o) Selten mit Kompression
- (-) Langsame out-of-order Inserts bzw. Updates
- (-) Keine Intervallarithmetik
- (-) Häufig enge Retention Policies

Datenmanagement und Betriebskosten

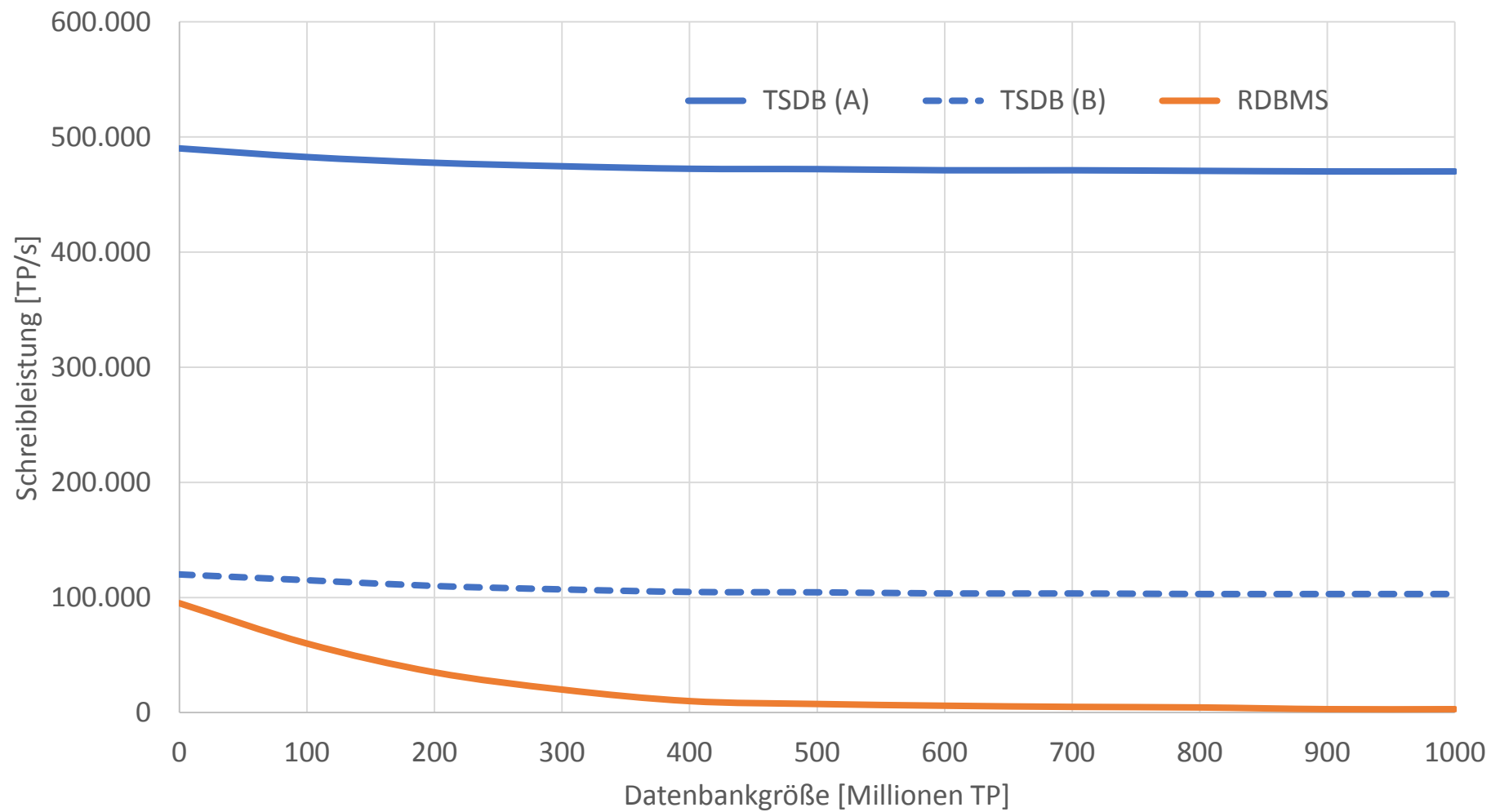




Schreibmuster

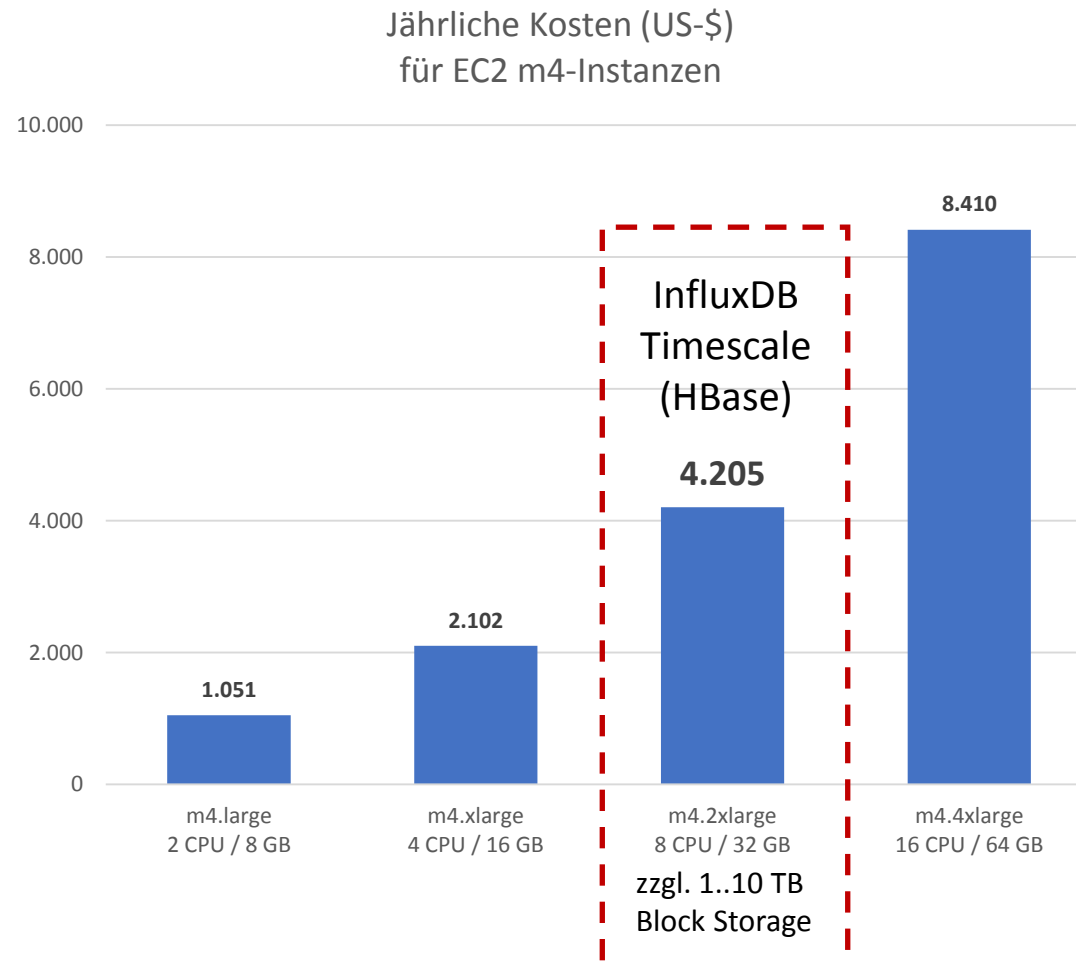
- Datenerfassung im Feld
 - MQTT QoS für die Nachrichtenübertragung:
At least once / at most once / exactly once
append >> update >> out-of-order insert
- Abgleich / Harvesting von Datenarchiven
 - Fortschreibung neuer Daten
 - Regelmäßige Aktualisierung „alter“ Daten
append ~ update >> random insert

Schreibperformance





Betriebskosten



Testkonfiguration

6 Mrd. Messwerte

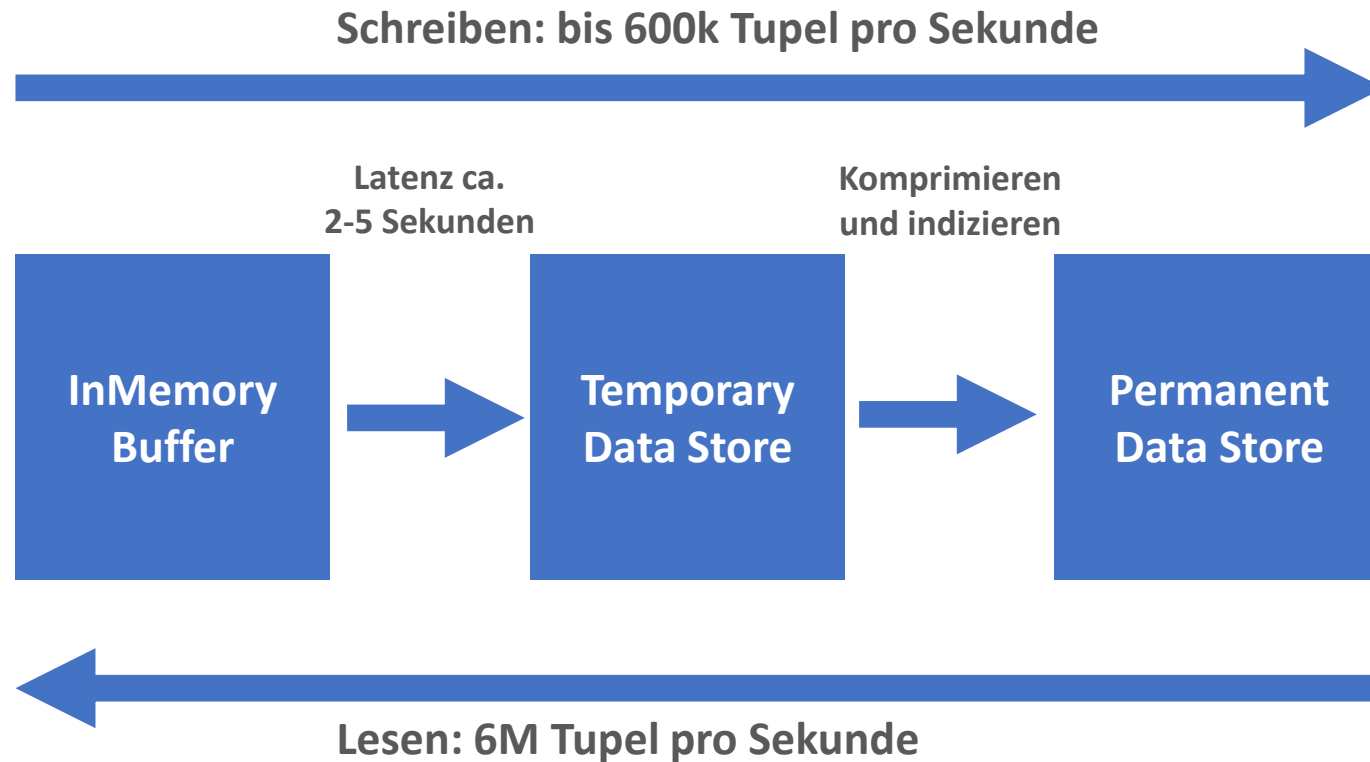
400k Sensoren

Datenvolumen:

~ 600 GB .. > 10TB



DB-Architektur *SensorHub*



Testkonfiguration

6 Mrd. Messwerte

400k Sensoren

Datenvolumen:

~ 5 GB komprimiert

CX31	CX41
€ 10,59 pro Monat	€ 18,92 pro Monat
€ 0,017 / Stunde	€ 0,031 / Stunde
2 vCPU 8 GB RAM 80 GB NVMe SSD 20 TB Traffic	4 vCPU 16 GB RAM 160 GB NVMe SSD 20 TB Traffic



Fazit und Ausblick

- Machbarkeitsbeweis erbracht: <http://opensensorweb.de>
- Anbindung weiterer Archive
- Zuschnitt domänenspezifischer Datenprodukte
- Unterstützung weiterer Datentypen
(Rasterdaten, klassische Geometriedaten)
- Klärung nicht-trivialer Datenlizenzen



PIKobytes
Königsbrücker Straße 124
01099 Dresden

www.pikobytes.de

