

Improving data quality by lowering usage barriers

ODQ2015 - Open Data Quality: from Theory to Practice

Matthew Fullerton ([@mattfullerton](https://twitter.com/mattfullerton)) | Florian Schimandl ([@fschimandl](https://twitter.com/fschimandl))



SMARTLANE - smartlane.de | OK Lab München - codefor.de/muenchen | OKF Germany - okfn.de

Open Data: Quantity is not quality
(catalogs are unhelpful?)



Transparenzportal Hamburg

transparenz.hamburg.de

bereitgestellt von

hamburg.de



NEUE SUCHE

TRANSPARENZGESETZ

OPEN DATA

HILFE

KONTAKT



Portal durchsuchen

nach Stichwort

SUCHEN

ERWEITERTE SUCHE

HILFE ZUR SUCHE

Sortierung nach

RELEVANZ

▼ DATUM

TITEL

1

2

3

...

1241

»

Ergebnis 1-20 - von **24813 Ergebnissen**. Treffer je Seite: 20 | 50 | 100



Das Baugewerbe in Hamburg im Dezember 2014

Informationsgegenstand: Statistiken und Tätigkeitsberichte

11.03.2015

XLSX

ZIP

...

IHRE SUCHANFRAGE

keine

SUCHERGEBNISSE EINSCHRÄNKEN:

KATEGORIEN

Bevölkerung (447)

Bildung & Wissenschaft (541)

Datenansicht

Daten von diesen
Organisationen auflisten

Offenheit

Unbekannt (1268)

Offen (36)

Nicht offen (33)

Quelle

Google (788)

Crawl (256)

Bing (242)

Data Catalog (33)

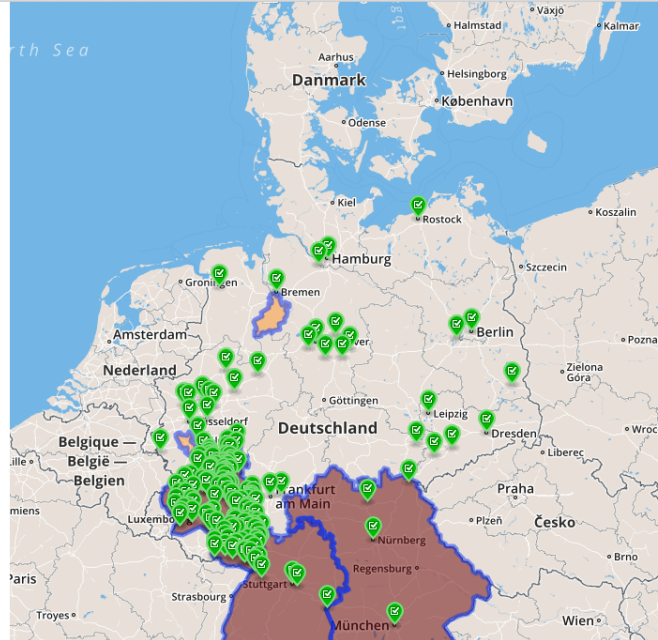
Manual (18)

Gruppen

Geographie, Geologi...
(3335)

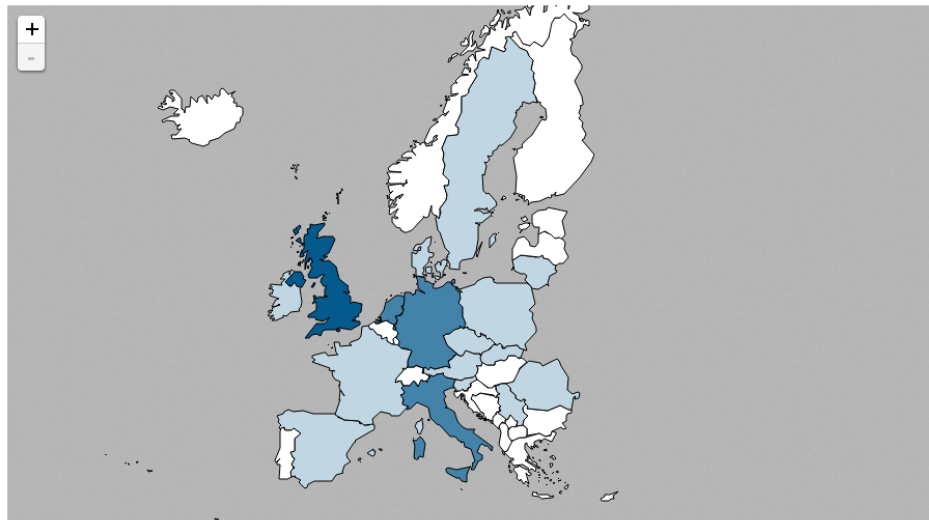
Bildung und Wissens...
(2267)

Infrastruktur, Baue... (652)





Availability map



This map shows the current availability for datasets across the European countries. Use the controls on the left or the mouse wheel to navigate, and click a country to see the number of packages available.

European
Data
Catalogues
Visualisations

- Dashboard
- Open licenses
- Total vs open
- Unique publishers

HARVESTED DATA CATALOGUES

70

From 226 Planned Catalogues

COUNTRIES HARVESTED

18

From 28 European Member Countries

OPEN LICENSES

23%

In 202,474 distributions with a license attached

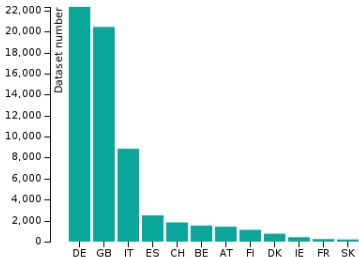
MACHINE-READABLE

17%

From 66,747 distributions harvested

DATASETS HARVESTED ⁰

163,823



TOTAL SIZE

0.52 TB

Distributions ⁰

Total distribution

381,735

Distribution maximum size

37.8 MB

Distributions with a license attached

202,474

53 %

Distributions with an open license

88,603

23 %

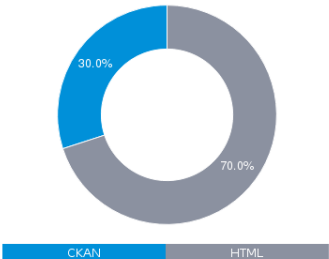
Distributions machine-readable

66,747

17 %

Software platform in the harvested catalogues

The OpenDataMonitor platform provides information and statistics on European open data catalogues, harvested through two methods. The first, CKAN, is currently the most widely used open source data management system. In addition, for those catalogues that are not built on an open source management system, HTML scraping procedures are being used to extract information from websites. We are also exploring the feasibility of working with other publishing platforms to improve the coverage of data.



European Commission awards the deployment of the pan-European Open Data Portal to a consortium led by Capgemini

February 5, 2015 | Press release

Also available in: [french](#) | [norwegian](#)

Share:

 PDF - 184.06 kB

Paris - Capgemini, one of the world's foremost providers of consulting, technology and outsourcing services, has announced that it has been awarded a three-year contract, worth several million Euros by the European Commission, for the deployment of an EU open data core platform. The goal of this project is to develop, launch and operate a single large-scale pan-European Open Data Portal, which will bring together public data resources from all over Europe in one portal covering all 39 European countries, improving the discoverability and the ability to re-use this data. The data is to be made fully public, so the information will be available to every European resident who has access to the internet.

Capgemini Consulting, the strategy and transformation consulting arm of the Capgemini Group, will lead this three-year engagement, co-ordinating a number of consortium members and subcontractors, including Capgemini's Group subsidiary - [Sogeti](#), [Intrasoft International](#), the [Open Data Institute](#), [Fraunhofer FOKUS](#), [con terra](#), the [University of Southampton](#) and [timelex](#).

But... metadata and exchange quality is important
(catalogs are good!)

Catalogmetadata

A website collecting the metadata (API calls or file downloads) offered by data catalogs in Germany

Aachen

[JSON](#)

Berlin

[JSON](#) (filtered from [govdata.de](#))

Bonn

- [JSON](#) (direct from API)
- [Extended JSON](#) (dump of calls to API for all datasets as well as some data from the overview JSON above)

Braunschweig

[XML](#)

Bremen

[JSON](#)

Landkreis Dinklage

Ulm

- [XML](#): simple concatenation of all files as offered by the city of Ulm and wrapped within one xml tag
- [JSON export](#) of much of the data from the XML, mostly flattened, and including data missing from the XML and scraped from the HTML pages (license, file uris)



- {

title: "Aktuelle Meldungen und Pressemitteilungen für das Jahr 2000",

name: "bremen236_c_8635_de",

notes: "Der Datensatz enthält einen Auszug aus dem Informationsregister: die aktuellen Meldungen und Pressemitteilung
enthält einen Auszug aus dem Informationsregister: die aktuellen Meldungen und Pressemitteilungen der Senatspressestelle
kompletter Export dieser Daten "

email: "datenerfahrungen.bremen.de"

}

],

- dates: [

- {

role: "erstellt",

date: "2012-12-10 15:43:30"

},

- {

role: "aktualisiert",

date: "2015-03-13 18:24:04"

},

- {

role: "veroeffentlicht",

date: "2000-01-01 00:00:00"

}

],

- terms_of_use: {

licence_id: "cc-by",

attribution_text: "Creative Commons Namensnennung (CC BY 3.0)",

licence_url: "",

other: ""

},

temporal_coverage_from: "2000-11-01 00:00:00",

temporal_coverage_to: "9999-01-01 00:00:00",

metadata_original: "",

metadata_original_portal: "",

ogdd_version: "v1.0.0",

temporal_granularity: "Tag"

```
[  
  - {  
    maintainerphone: "07311611037",  
    distributorfax: "0731161808882",  
    maintainerorganisation: "Stadt Ulm",  
    abstract: "Die Wohnbevölkerung der Stadt Ulm nach Altersjahren, Staatsangehörigkeit und Geschlecht enthä  
in einer Zahl zusammen gefasst. ",  
    distributorphone: "07311618882",  
    accessconstraints: "keine",  
    maintainername: "Lucia Erdt",  
    distributoremail: "statistik@ulm.de",  
    - keywords: {  
      GOV-Data Kategorien: "Bevölkerung",  
      Klassifikation GDI BW: "Statistik",  
      GEMET - INSPIRE themes, version 1.0: "Verteilung der Bevölkerung - Demografie"  
    },  
    frequpdated: "annually",  
    datacatalogdate: "2013-12-10",  
    accessconstraints: "",  
    contactphone: "07311618882",  
    title: "Wohnbevölkerung nach Altersjahren, Staatsangehörigkeit und Geschlecht der Stadt Ulm",  
    maintenanceinfo: "annually",  
    contactemail: "statistik@ulm.de",  
    contactname: "Winfried Baumeister",  
    maintaineraddress: "Frauenstraße 19 Ulm 89073 Germany",  
    distributorrole: "Point of Contact",  
    maintainerfax: "0731161808777",  
    datadate: "2011-12-31T00:00:00",  
  }
```

**Stadtrats-Übersicht**

Schnellsuche

Dokumentensuche

StR-Anträge/Anfragen

Offene StR-Anträge/Anfragen

StR-Vorlagen/Beschlüsse

StR-Sitzungen

Sitzungskalender

StR-Gremien / Ausschüsse

Fraktionen

Stadratsmitglieder

Referenten

Hilfe / Glossar

Meinung / Probleme

Barrierefreiheit

Impressum

Bezirksausschuss-Übersicht

RIS - RatsInformationsSystem - Stadtrat

Was macht der Münchner Stadtrat?

Darüber informiert Sie **RIS**, das **RatsInformationsSystem** der Stadt München. RIS stellt Ihnen die Anträge, Anfragen sowie die öffentlichen Vorlagen und Beschlüsse zur Verfügung, mit denen sich der Münchner Stadtrat befasst.

Wer im Stadtrat sitzt, welche Ausschüsse es gibt und vieles mehr erfahren Sie hier.



Schnellsuche / Dokumentensuche

Mit den Suchmöglichkeiten finden Sie alle im **RIS - RatsInformationsSystem** für den Bereich "Stadtrat" eingestellten öffentlichen Informationen



StR-Anträge / Anfragen

Dies ist eine Übersicht der vom Münchner Stadtrat in den **letzten Tagen** gestellten Anträge und Anfragen



StR-Sitzungen

Hier erhalten Sie eine Übersicht der Sitzungstermine aus den **vergangenen** Tagen (StR-Gremien / Ausschüsse)



StR-Vorlagen / Beschlüsse

Hier erhalten Sie eine Übersicht der in den **letzten Tagen** gefassten Beschlüsse (inkl. Vorlagen)



Sitzungskalender

Hier erhalten Sie eine Übersicht der in den **nächsten Wochen** vorgesehenen Sitzungstermine (StR-Gremien / Ausschüsse)



Fraktionen / Gruppierungen

Hier finden Sie eine Übersicht aller Fraktionen und Gruppierungen der **derzeitigen** Wahlperiode



StR-Gremien / Ausschüsse

Hier erhalten Sie eine Übersicht aller **aktuellen** Stadtratsgremien und Ausschüsse



Stadratsmitglieder / Referenten

Hier erhalten Sie eine Übersicht **aller amtierenden** Münchner Stadratsmitglieder

Quality in practice:

Quality is domain specific.

Quality (measurement) is hard.

Quality indicators for traffic data, e.g.:

- Availability
- Completeness/Coverage
- **Currentness/Timeliness**
- Correctness
- ...

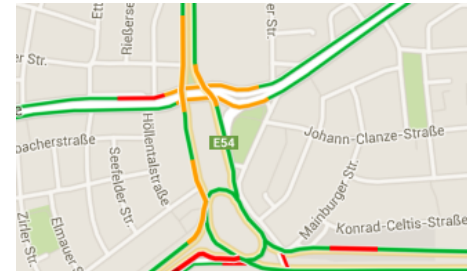
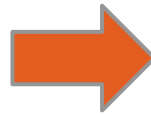
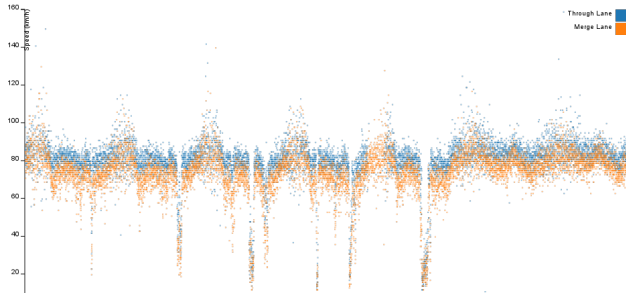
Necessity strongly depending on Application!

Example:

- **Real-Time** System for Traffic State Analysis/Estimation
- Traffic-Message-Channel (TMC) - System
- ...

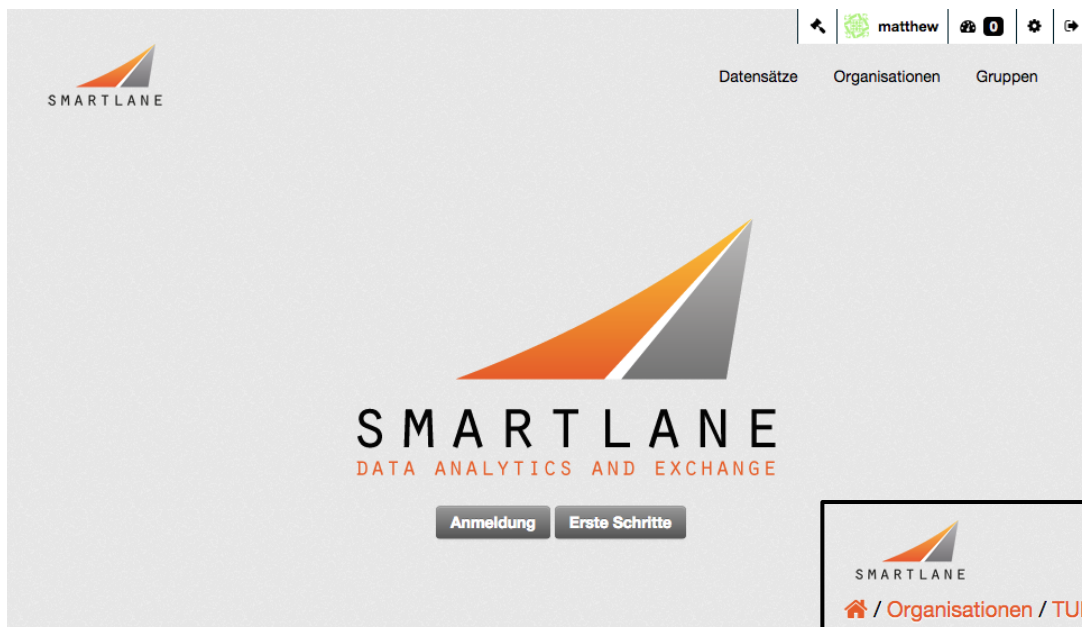
BEWARE! Garbage-In → Garbage-Out

Examples: Broken detectors, Long transmission delays, no detection of traffic jam dissolution with taxi data (knowledge dissemination among drivers), ...



Source: Google

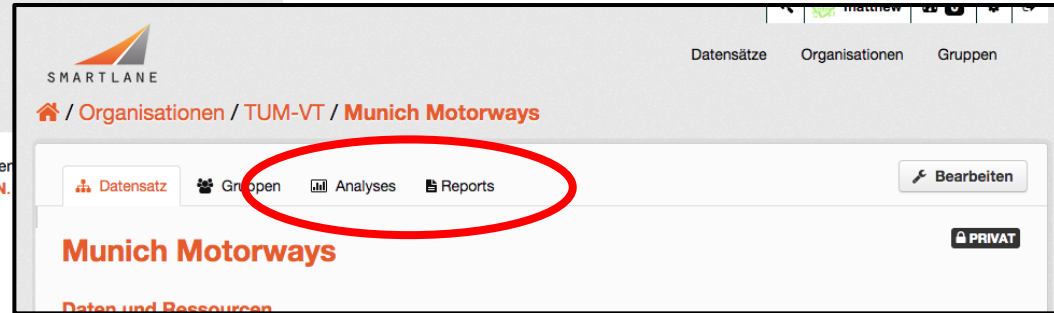
Quality: Tooling for measurement and improvement



- Data management
- Data **quality assessment**
- Data analysis
- **Workflow** management
- Visualization
- Reporting

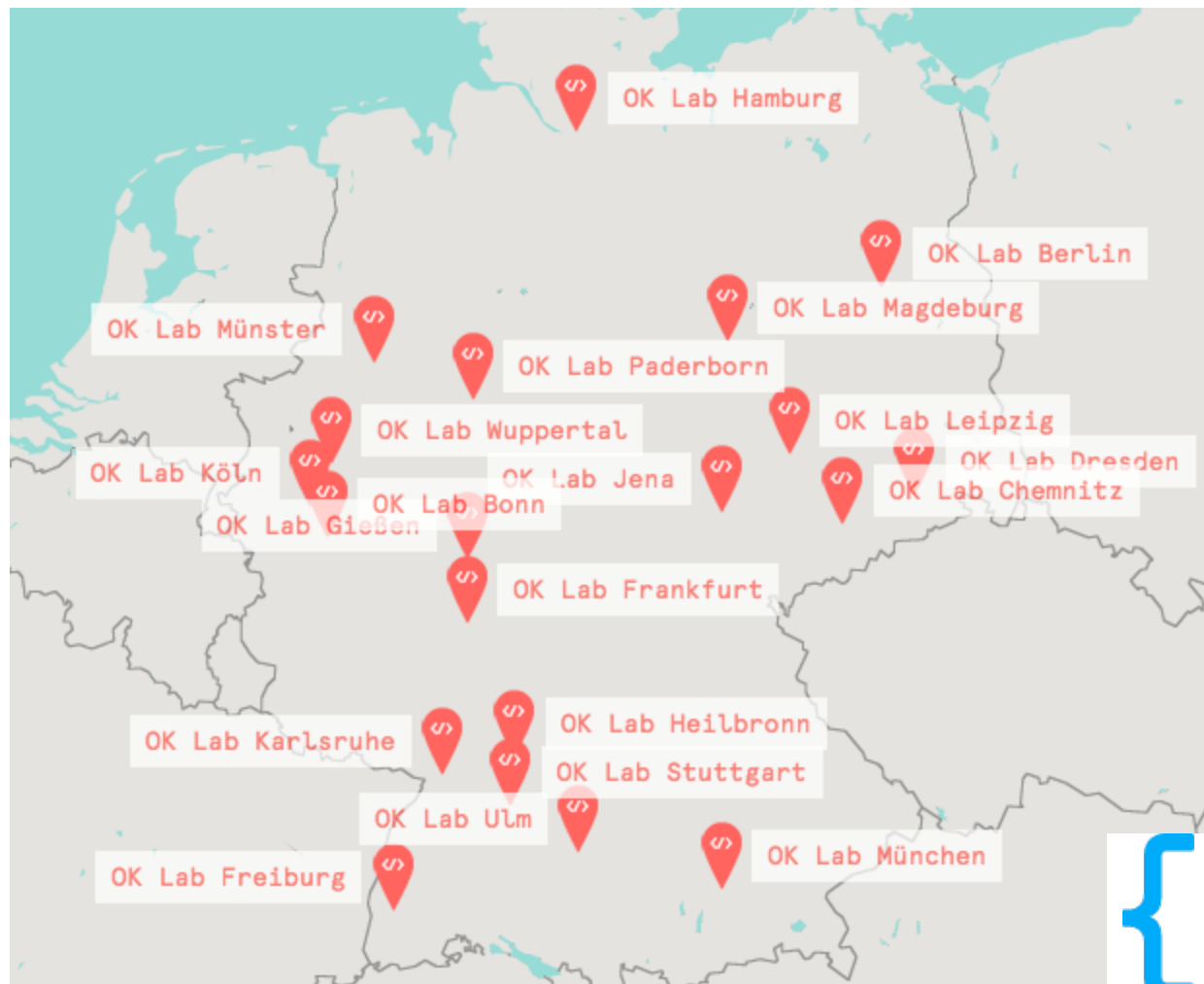
Sprache: Deutsch

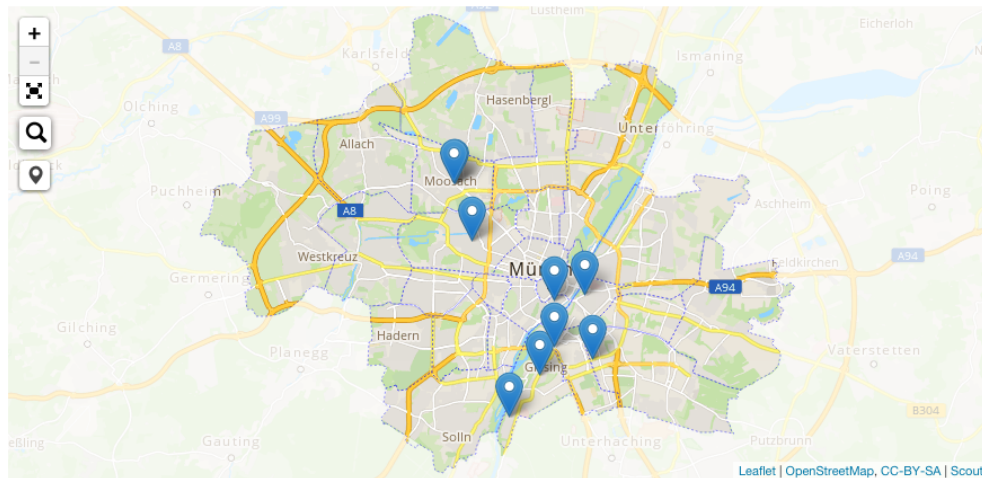
SMARTLANE wird innerhalb des EXIST-Gründungsvorhabens
Datenplattform basiert auf dem Open Source Projekt **CKAN**.



- Funded through the German Government 'EXIST' Stipend scheme until 12/2015
- Functionality exposed via CKAN extension
- Available to test users Q3 2015

Citizens improve quality



**SO FUNKTIONIERT STADTPOLITIK**

KOMMUNALPOLITIK IN MÜNCHEN EINFACH ERKLÄRT

**TERMINE**

WANN FINDEN STADTRATS- / AUSSCHUSSSITZUNGEN STATT?

**THEMEN**

DOKUMENTE, GEGLIEDERT NACH THEMA UND REFERAT

**DOKUMENTENSUCHE**

DURCHSUCHE 157.405 DOKUMENTE / 613.459 SEITEN

**PERSONEN**

WER SITZT IM STADTRAT / IN DEN BEZIRKSAUSSCHÜSSEN?

**E-MAIL-BENACHRICHTIGUNG**

PER MAIL ÜBER NEUE DOKUMENTE INFORMIERT WERDEN

**ÜBER MÜNCHEN-TRANSPARENT**

ÜBER DIESE SEITE

**for**
München

Quality starts with quantity!?

Improving data quality by lowering usage barriers

ODQ2015 - Open Data Quality: from Theory to Practice

Matthew Fullerton ([@mattfullerton](https://twitter.com/mattfullerton)) | Florian Schimandl ([@fschimandl](https://twitter.com/fschimandl))



SMARTLANE - smartlane.de | OK Lab München - codefor.de/muenchen | OKF Germany - okfn.de
