

# Open Data Zensus

**Automatisches Ranking der Open Data  
Programme deutscher Städte**

**Wer ist Nummer 1 in Deutschland?**

**Rostock!!!**

Das ist eine schöne Motivation für uns sich für die  
Bereitstellung "Offener  
Daten" in der Verwaltung der Hansestadt Rostock so intensiv  
zu engagieren.

München ist 2. :)

# Ein paar Zahlen

- > 26.000 Datensätze
- 146 Städte
- 53 Harvester
- täglich aktualisiert

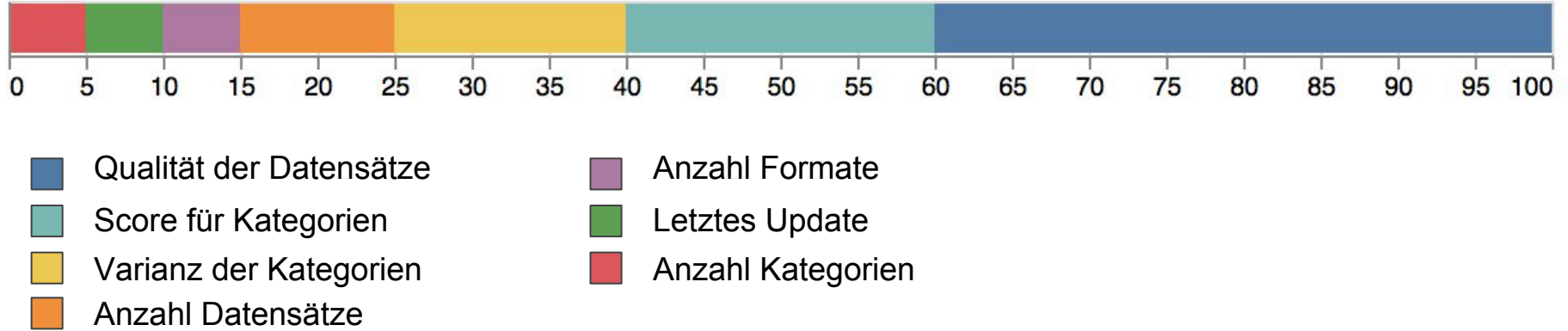
— — —

# Was will ich mit dem Ranking erreichen?

- Benchmark der Qualität
- Keinen Mehraufwand erzeugen
- Fortschritt zeigen
- Diskussionsanstoß



# Scoring





# Kriterium #1: Qualität der Daten

# Qualität der Daten

Gewicht: 40%

1. Lizenz des Datensatzes
2. Format des Datensatzes
3. Letztes Update

Max 3 Punkte

Arithmetisches Mittel  
aller Datensätze ergibt  
die Punktzahl

---

# Score für Kategorie

Gewicht: 20%

1 Punkt: Datensätze in  
allen Kategorien

0,5 Punkte für Datensätze  
in mind. der Hälfte der  
Kategorien

sonst 0 Punkte

---

# Score für Varianz

Gewicht: 15%

viele Datensätze in  
wenigen Kategorien →  
weniger Punkte

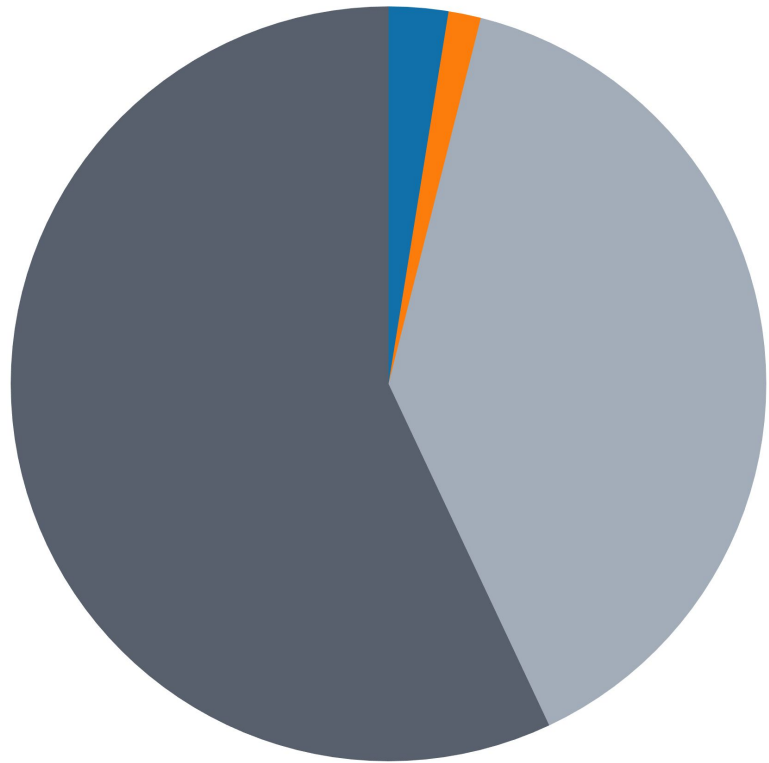
gleichverteilte Daten →  
mehr Punkte

---

# Beispiel: Varianz

Viele Datensätze auf  
wenige Kategorien verteilt

→ Varianz aufnehmen und  
15% Gewichtung



# Score für Datensätze

Gewicht: 10%

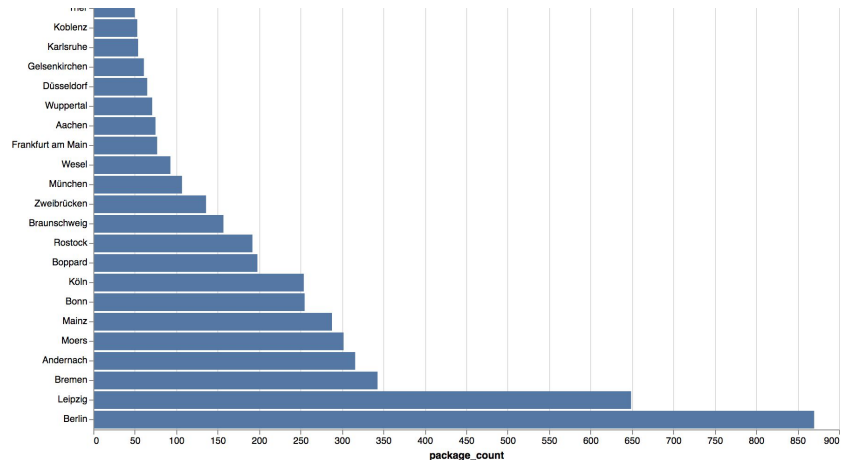
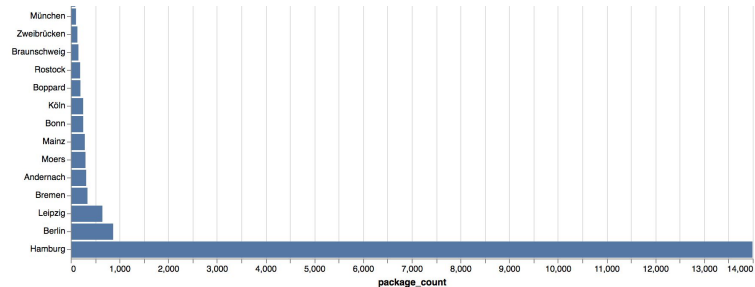
viele Datensätze → mehr  
Punkte

---

# Beispiel Datensätze

Stadtstaaten haben  
deutlich mehr Daten

→ weniger Gewicht der  
reinen Anzahl



# Score für Formate

Gewicht: 5%

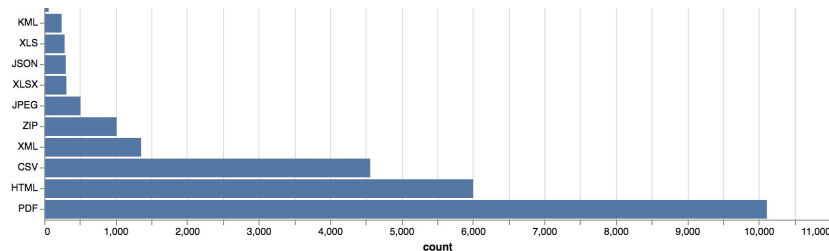
viele Formate → mehr  
Punkte

---



# Beispiel: Formate

→ Focus auf Offene und  
maschinenlesbare Formate



# Score für letztes Update

Gewicht: 5%

Hohe Punktzahl je  
aktueller der letzte  
Datensatz ist

---

# Score für Formate

Gewicht: 5%

Je mehr Formate umso höher  
die Punktzahl

---

**Was haben die Top 5 gemein?**

**fast 100% offene Lizenzen**

**$\frac{3}{4}$  offene und maschinelesbare Formate**

**Datensätze in fast allen Kategorien**

**Gut gepflegte Metadaten**



# Gute Beispiele für Datensätze

- Verschiedene Datenformate in einem Datensatz
- Verschiedene Jahre als ein Datensatz mit mehreren Ressourcen
- Ressourcen direkt verlinkt
- 

---

# Schlechtere Beispiele

- jeden einzelnen Abschnitt eines BPlans als Datensatz
- Externer Downloadlink
- Jedes Jahr als neuen Datensatz

**Datenquelle: [offenedaten.de](https://offenedaten.de)**

# Warum offenedaten.de?

- nicht alle Städte auf govdata.de
- mehr Kontrolle über die Metadaten

Ziel:

- Alle Städte auf govdata zu bringen  
(Verantwortung in den Ländern)
- Einheitliche Metadaten

— — —

# Bekannte Probleme

- unterschiedliche Metadaten in den Portalen
- Update Zyklus zu kurz?
- Scoring
- Neue Städte müssen per Hand eingetragen werden
- Löschen von Datensätzen grundsätzlich ein Problem

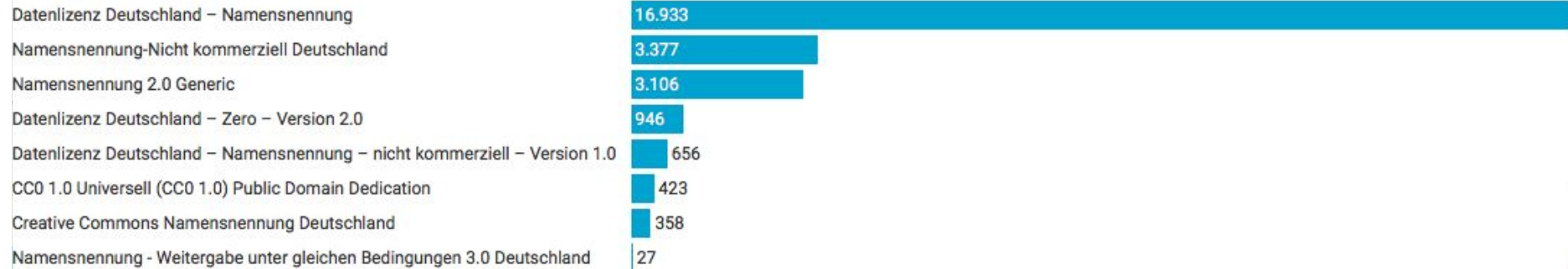


# Datenportale in D



[Get the data](#) • Created with Datawrapper

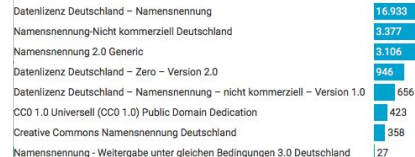
# Lizenzen in D



[Get the data](#) • [Created with Datawrapper](#)

# Lizenzen in D

## Lizenzen in Deutschland



[Get the data](#) • Created with [Datawrapper](#)

Lizenz	Anzahl Datensätze
Datenlizenz Deutschland - Namensnennung	16.933
Namensnennung - Nicht Kommerziell	3377
Namensnennung	3106
Datenlizenz Deutschland - Zero	946

— — —



Wie geht es weiter?

# Roadmap

— — —

- Feedback zum Scoring
- Interviews mit Städten
- Ranking der einzelnen Kategorien
- Umstellen auf govdata
- Validierung von XLS und CSV Dateien
- Unterscheidung von Landes und Kommunaldaten
- Ranking nach Stadt/Gemeinde/Kreise
- OGC Webservices mit aufnehmen (WMS, WFS, etc.)
- ...

**Weiterfinanzierung: leider offen**

**Her mit euren Fragen!**