Simon Jockers @sjockers

Hacking the statistics office

github.com/datenguide



datengui.de @datenguide

Official Statistics

- Statistics published by governments and other public bodies
- Statistics about society, the environment or the economy
- Statistics created by statistical
 offices on a national or regional level

Examples

Schools

(students, graduates by school type)

• Garbage removal (recycling, residual waste, etc.)

• **Migration & naturalisation** (origin, number of years in the country)



Image credits: Crop of original painting "Anbetung der Hirten" by Giorgione (public domain)

Where does data come from?

- Macro census ("Volkszählung")
- Micro census
- Secondary statistics
- Statistics based on population registries ("Registerzensus")

public good

-> open data

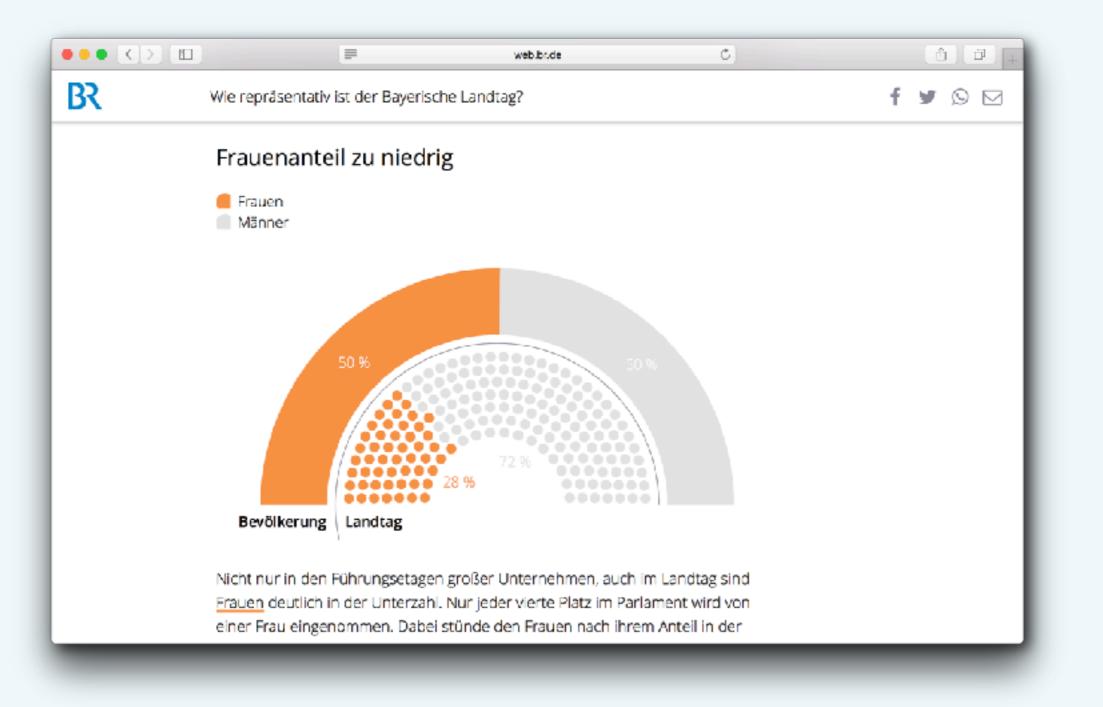
What is it good for?

- **Reporting**: Backing your story with numbers, finding stories in data
- Fact checking & debunking misinformation
- Lobbying for social causes

Limits

- Privacy: All data is **anonymised**
- Only aggregate-level data gets published (how many instead of who)
- People or companies cannot be identified

Example



web.br.de/interaktiv/landtag-demografie/

open data

-> available online, free for you to reuse.

5 🛧 Open Data

- 🔆 Available on the **Web**, under an **open license**
- **Structured** data (Spreadsheets over PDF)
- 🔆 Non-proprietary data formats (CSV over Excel)
- 🔆 Use **URIs** (let people link to data)
- ☆ Link to related data (discoverability, context)
 - (Tim Berners-Lee, <u>https://5stardata.info</u>)

Official statistics in Germany

- "Genesis" Data Portal Software
 - Several instances (genesis-online.de, regionalstatistik.de)
 - Data cube with SOAP-API
 - Frontend provides non-machine-readable
 CSV & Excel downloads
- Data Licence Germany Attribution Version 2.0 (dl-de/by-2-0)

Problems

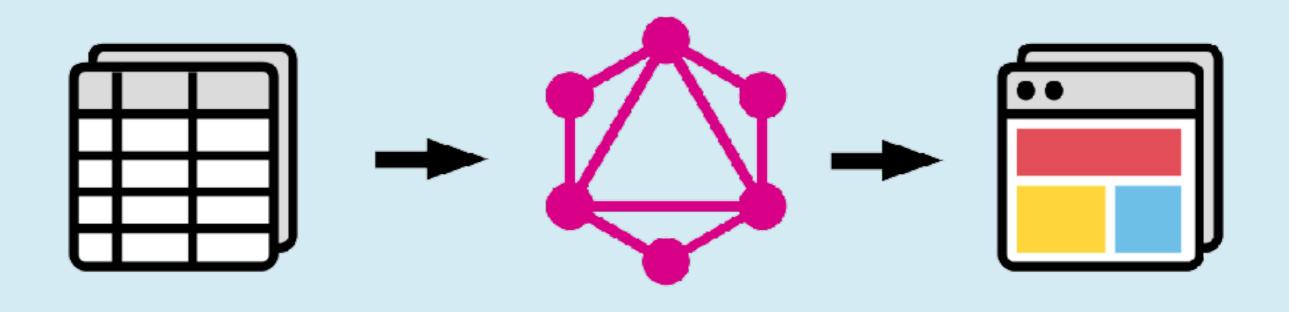
- Crappy software made for bureaucrats by bureaucrats
- Not usable for normal people such as journalists, civil society orgs
- Data is not easily reusable no open, machine-readable data formats; no modern APIs

open source

github.com/datenguide

$\bullet \bullet < >$		🗎 GitHub, Inc.			o.com/datenguide	Ċ	
O	Features Business	Explore	Marketplace	Pricing	This organization	Search	Sign in or Sign up
	dateng						
	repositories	ple 2	Projects O			Type: All	- Language: All -
daten Backend	guide-backend					Top languages	

Datenguide data pipeline



Data import

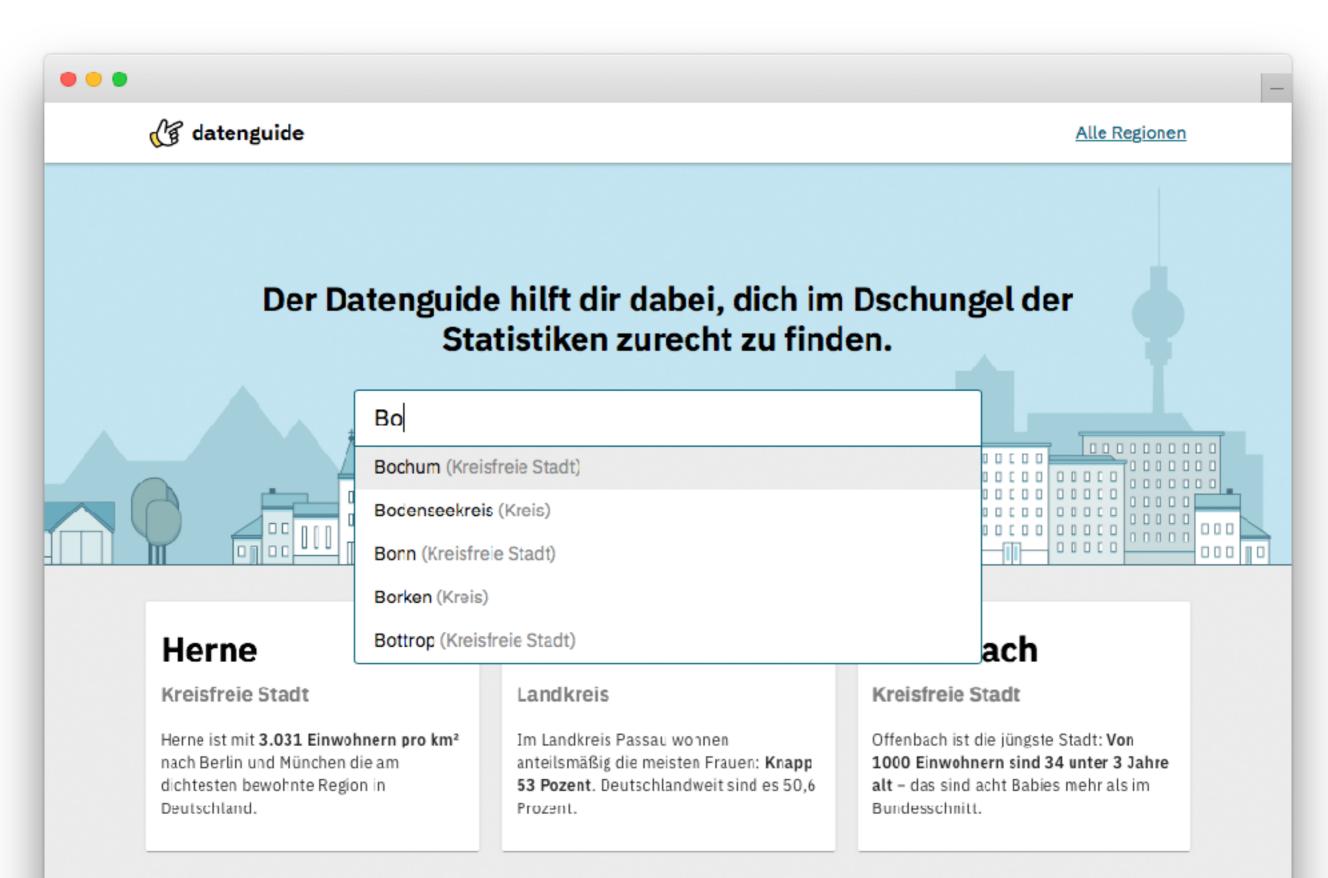
GraphQL API

Data portal

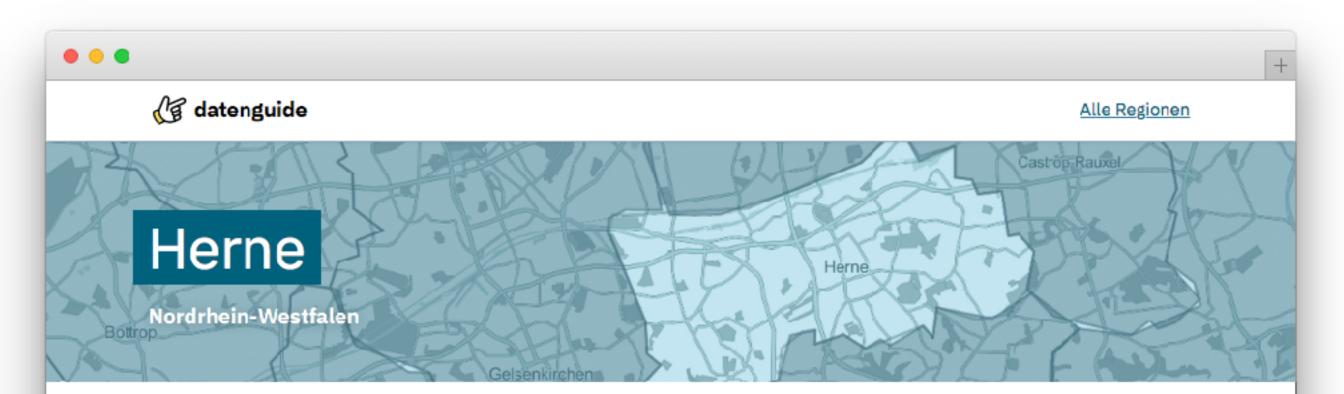
datenguide-frontend

- Generates individual pages for all
 German regions (NUTS 1-3):
 - 16 Federal states
 - ~ 400 Districts (Landkreise) and cities
- Visualises and explains the data
- Provides machine-readable CSV

datengui.de (frontend)

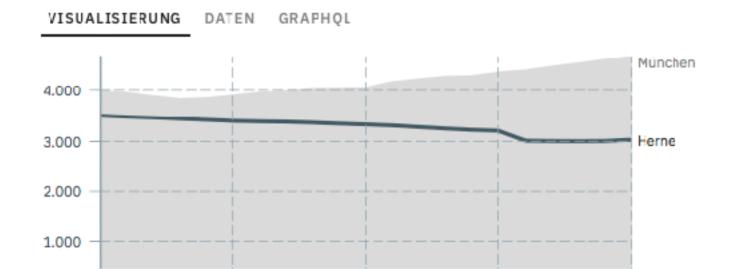


datengui.de (frontend)



Bevölkerungsdichte (Einwohner pro km²)

Die Bevölkerungsdichte ist mittlere Anzahl der Einwohner pro Quadratkilometer für ein bestimmtes Gebiet. Sie kann errechnet werden, indem man die Einwohnerzahl durch die Fläche des Gebietes teilt. In dieser Grafik vergleichen wir Herne mit München, der Stadt mit der höchsten Bevölkerungsdichte in Deutschland.



Einwohnerzahl 2017

155.851

🛱 51.42 km² Fläche

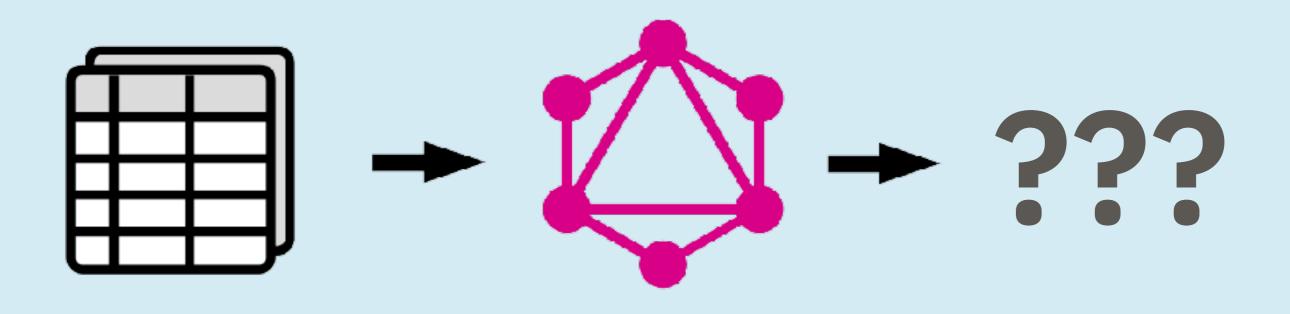
- 🚜 3.031 Einwohner pro km²
- 🖞 51% weiblich / 49% männlich

datenguide-frontend

- "JAM-Stack" (with Gatsby & React):
 - JavaScript for dynamic functionality
 - Built upon reusable APIs, accessed over HTTPS with JavaScript
 - Static markup, prebuilt at deploy time

github.com/datenguide/datenguide-frontend

Datenguide API



Data import GraphQL API Your use case

https://api.genesapi.org/ (alpha!)

datenguide-backend

- Python, Flask, and Elastic Search
- Imports official statistics data from "Genesis" instances, building upon github.com/pudo/regenesis
- Focus on data from **regionalstatistik.de**
- Provides a GraphQL API

github.com/datenguide/datenguide-backend

API demo

query { regions(nuts: 3, parent: "10") { name WAHLSR(year: "2017") { value

https://api.genesapi.org/ (alpha!)

Funded by:

Medieninnovationszentrum Babelsberg



Bundesministerium für Bildung und Forschung



fithub.com/datenguide

••• <>	🗎 GitHub, Inc. github.com/datenguide	Ċ +
🌍 Features Business Exp	ore Marketplace Pricing This organization	on Search Sign in or Sign up
Catenguid	Projects 0	
Search repositories		Type: All - Language: All -
datenguide-backend Backend for datenguide application		Top languages