

**Simon Jockers**

**@sjockers**

**Hacking the  
statistics office**

**👉 [github.com/datenguide](https://github.com/datenguide)**

 **datenguide**

**datengui.de**  
**@datenguide**

# Official Statistics

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

# Examples

- **Schools**  
(students, graduates by school type)
- **Commuters**  
(commuting from and to a city)
- **Garbage removal**  
(recycling, residual waste, etc.)



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

# Where does data come from?

- **Full population census** (“Volkszählung”)
- **Registry-based census**  
 (“Registerzensus” – coming in 2021)
- **Sampling** (“Makrozensus”)
- **Sector-specific censuses**  
 (e.g. agricultural census)

# 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)
- Neither people nor companies or other institutions can be identified

# **open data**

**-> available online,  
free for you to reuse.**

# Official statistics in Germany

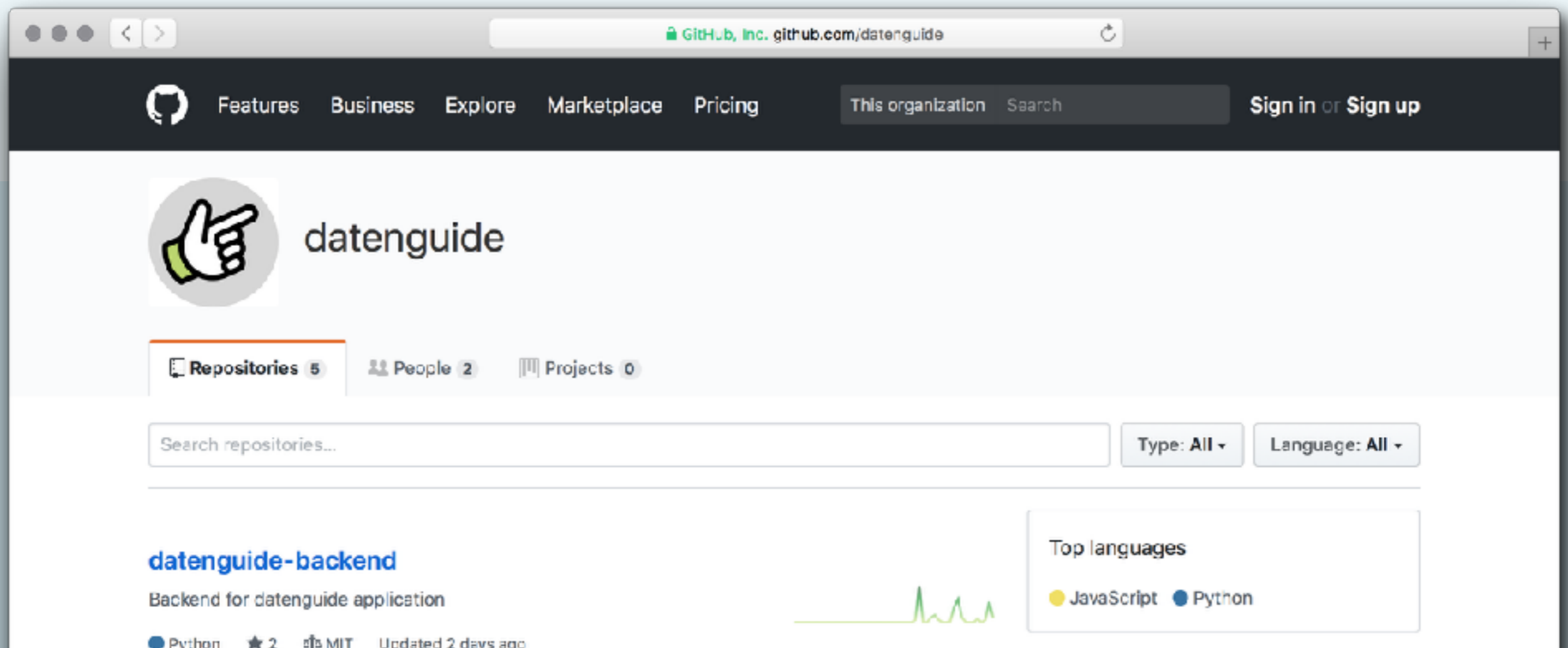
- **“Genesis”** Data Portal Software
  - **Several instances** (genesis-online.de, regionalstatistik.de)
  - Data cube with **SOAP-API**
  - Frontend provides non-machine-readable **CSV & Excel** downloads
- **Data Licence Germany** – Attribution – Version 2.0 (dl-de/by-2-0)

# Problems

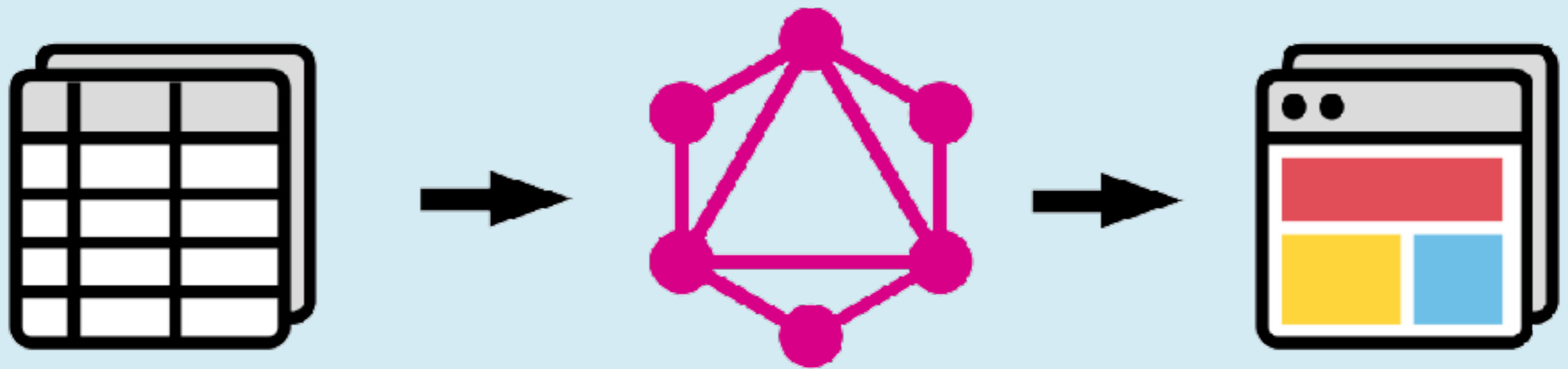
- **Crappy software**  
made for bureaucrats by bureaucrats
- **Not usable for normal people**  
such as journalists, civil society orgs
- **Data is not easily reusable**  
no open, machine-readable data formats;  
no modern APIs

# open source

👉 [github.com/datenguide](https://github.com/datenguide)



# Datenguide data pipeline



**Data import**

**GraphQL API**

**Data portal**

# How to use it

- API for downloading statistics as **machine-readable CSV and JSON**
- **GraphQL-API** for accessing statistics and meta data (early beta)
- **Data portal** – a web application for browsing data (coming soon)

# Datenguide in R & Python



**DatenguideR**

[github.com/CorrelAid/datenguideR](https://github.com/CorrelAid/datenguideR)



**datenguide-python**

[github.com/CorrelAid/datenguide-python](https://github.com/CorrelAid/datenguide-python)