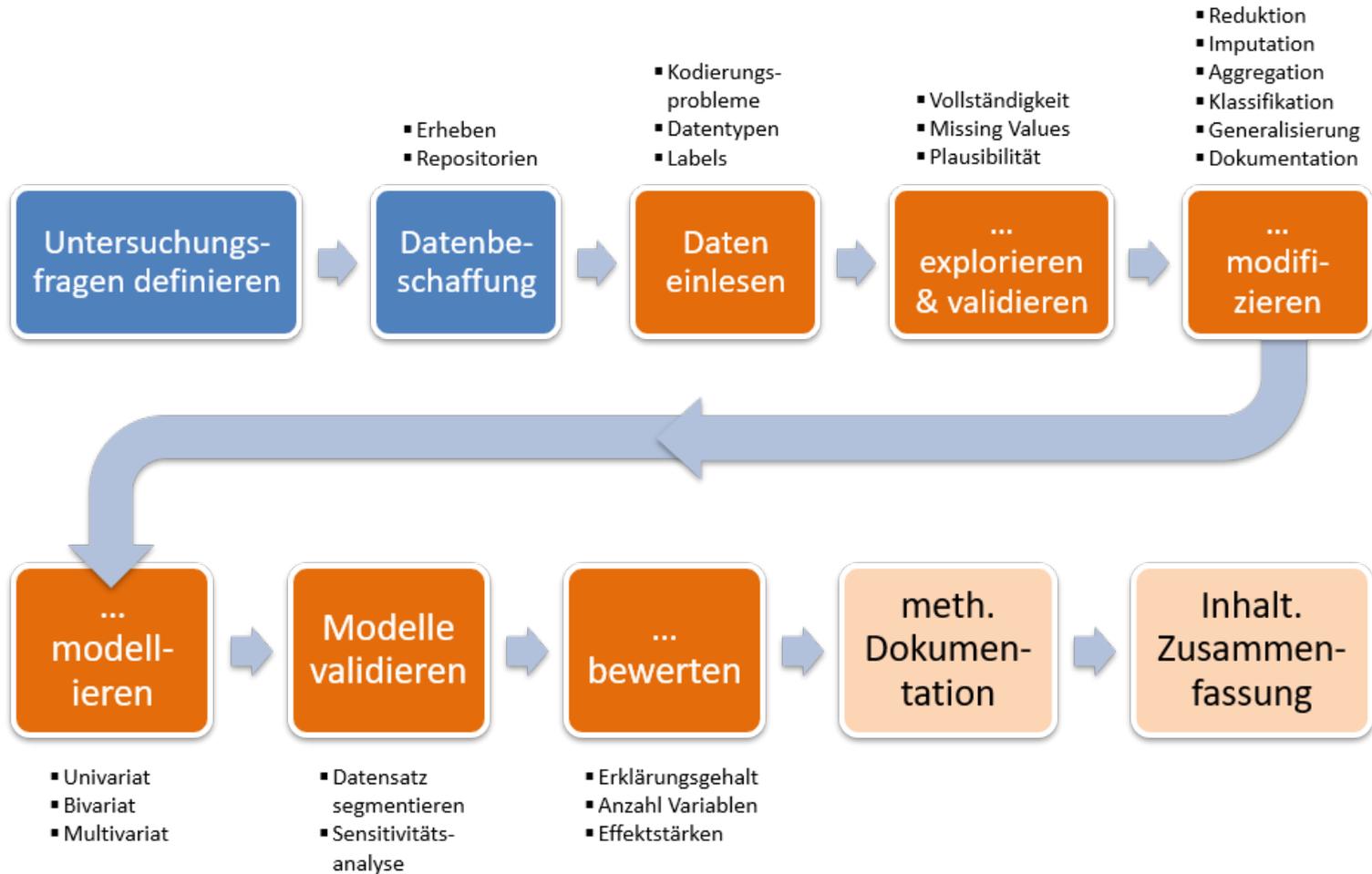


Warmup 2b: Auffindbarkeit & Nutzbarkeit von Daten

716408 | How 2 do Things with even more Numbers

KMH
WS 21-22 (updated: 2021-11-05)

Ein idealtypischer Ablauf





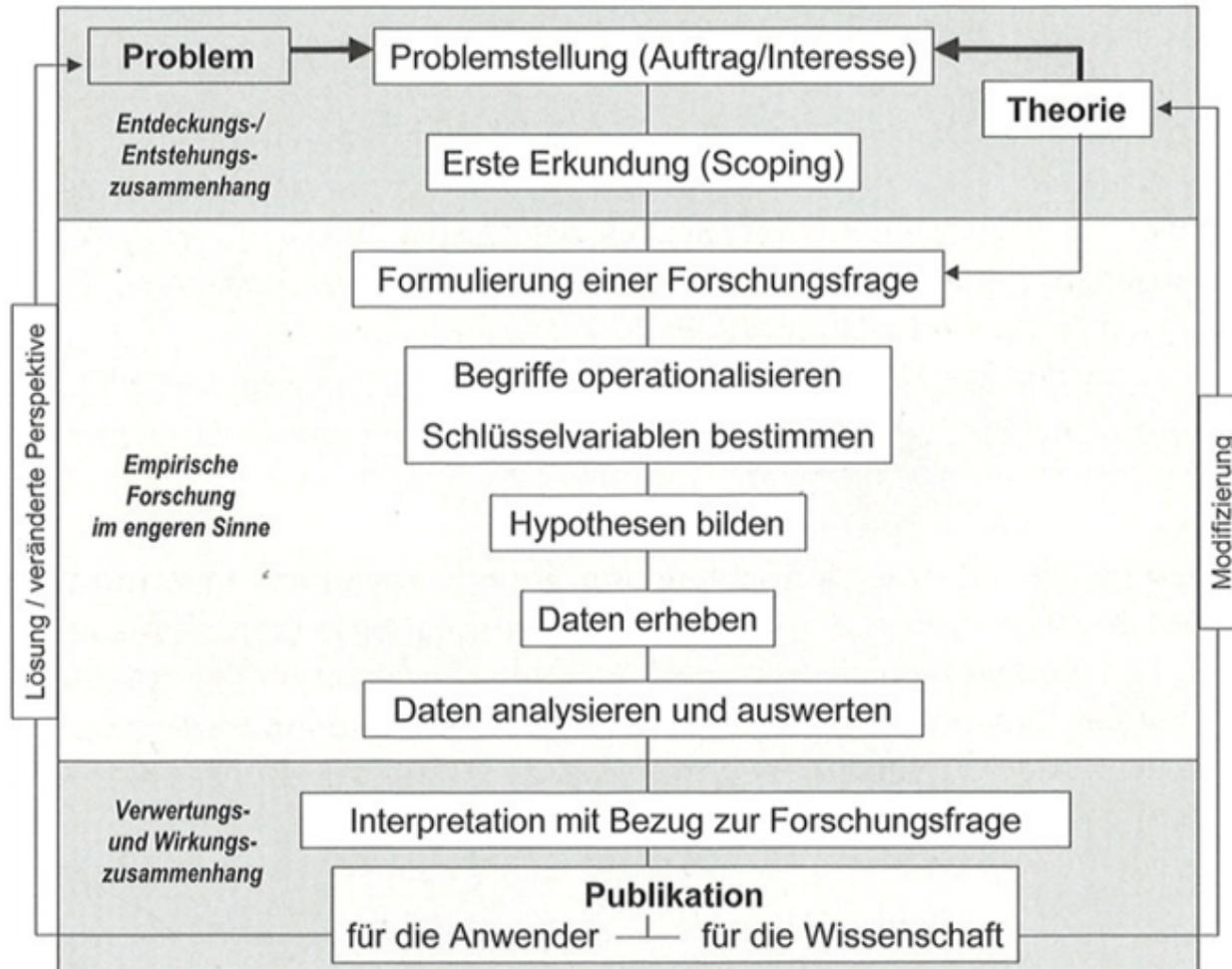
Woher kommen eigentlich die Daten?

2. Versuch: Wooooooooooooo!



(Giphy)

Woher kommen die Daten?



Exkurs: Die Primärerhebung quantitativer Daten

3 zentrale Methoden:

- **Beobachtung:**
systematisches Beobachten & quant. Registrieren „relevanter“ Sachverhalte
- **Nonreaktive Verfahren:**
quant. Erfassen einzelner Merkmale von Interesse in Texten, Bildern & Videos
- **Befragen:**
systematisch gesteuerte Kommunikation zwischen Personen; basierend auf standardisierten Fragen- & Antwortenkatalog

Aus Primär- werden Sekundärdaten

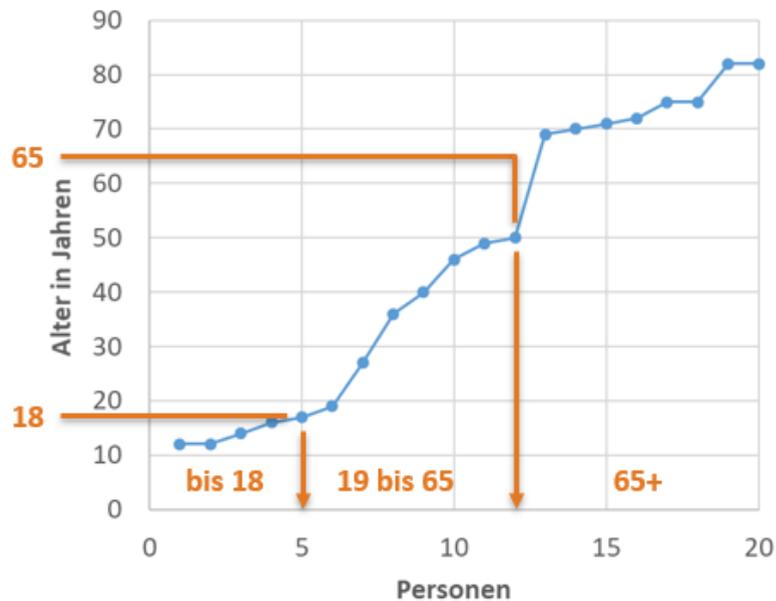
Sekundärdaten

durch Verarbeitung aus Primärdaten gewonnen

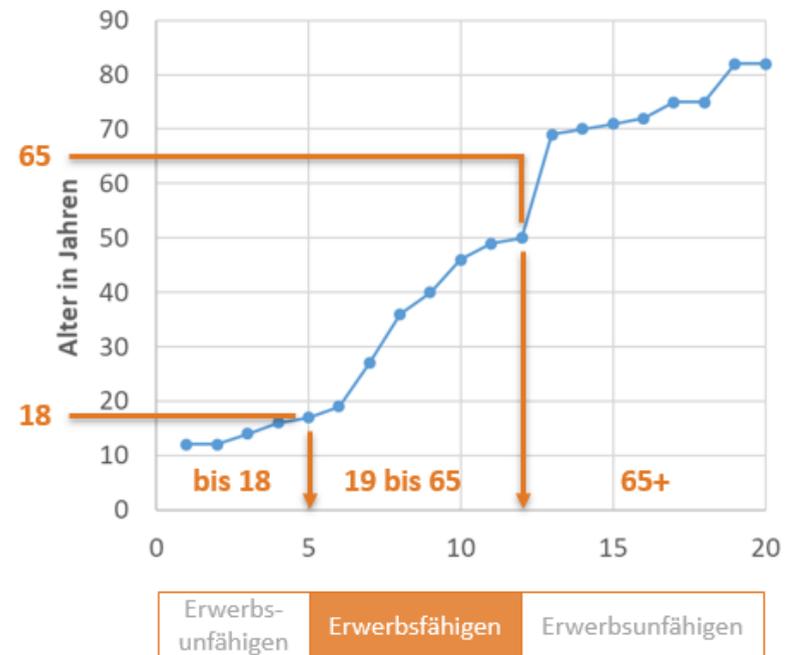
- Eigene Erstellung
- Erstellung durch Dritte
- @ Verarbeitung:
 - Klassifikation: Zordnung zu Klasse von Dingen mit gemeinsamen Eigenschaften
 - Aggregation: Gruppierung bestehender Klassen zu neuen (→ Definition einer neuen Klasse)
 - Generalisierung: Definition Teilmengen-Beziehung zwischen Elementen verschiedener Klassen

@ Klassifikation, Aggregation & Generalisierung

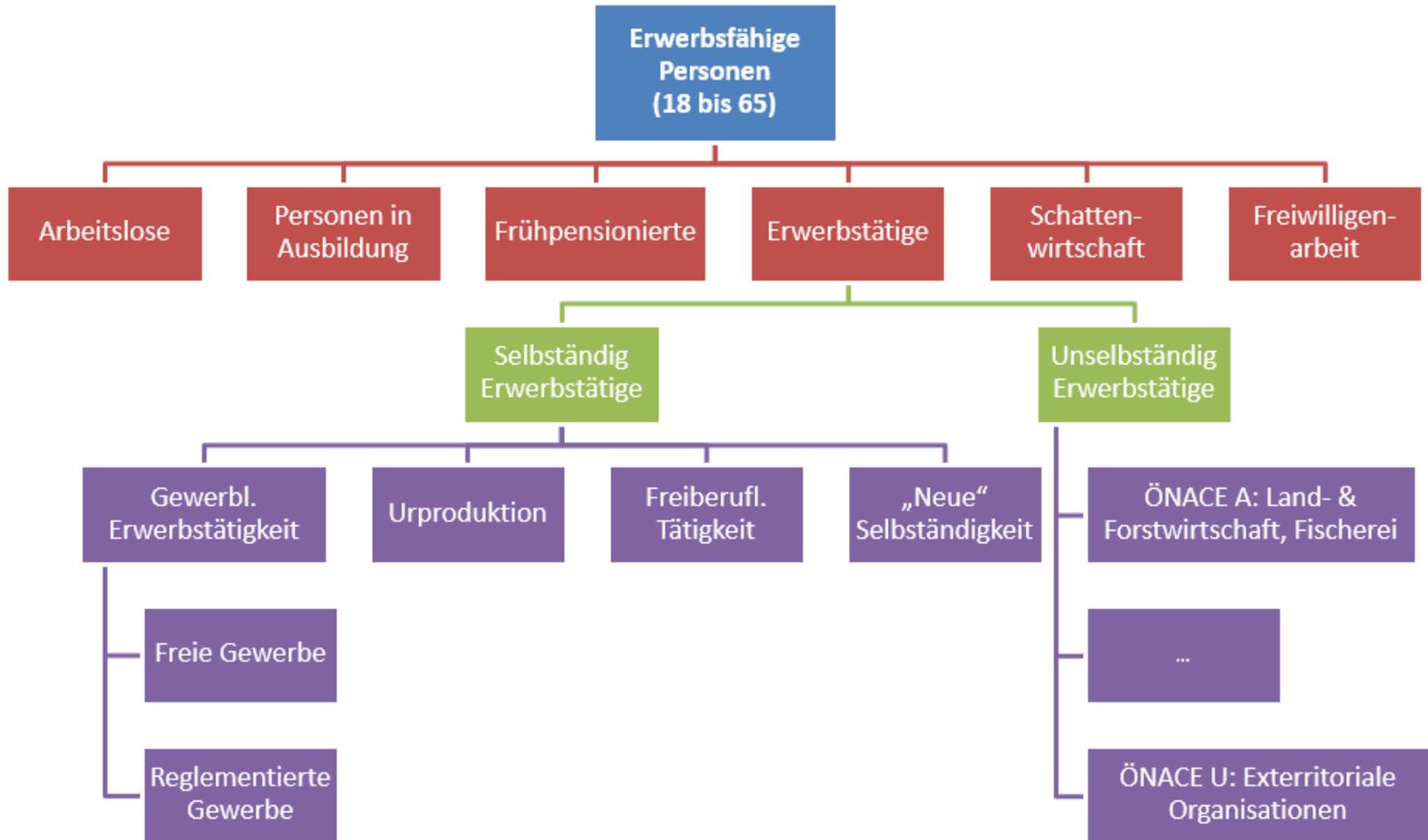
Klassifikation



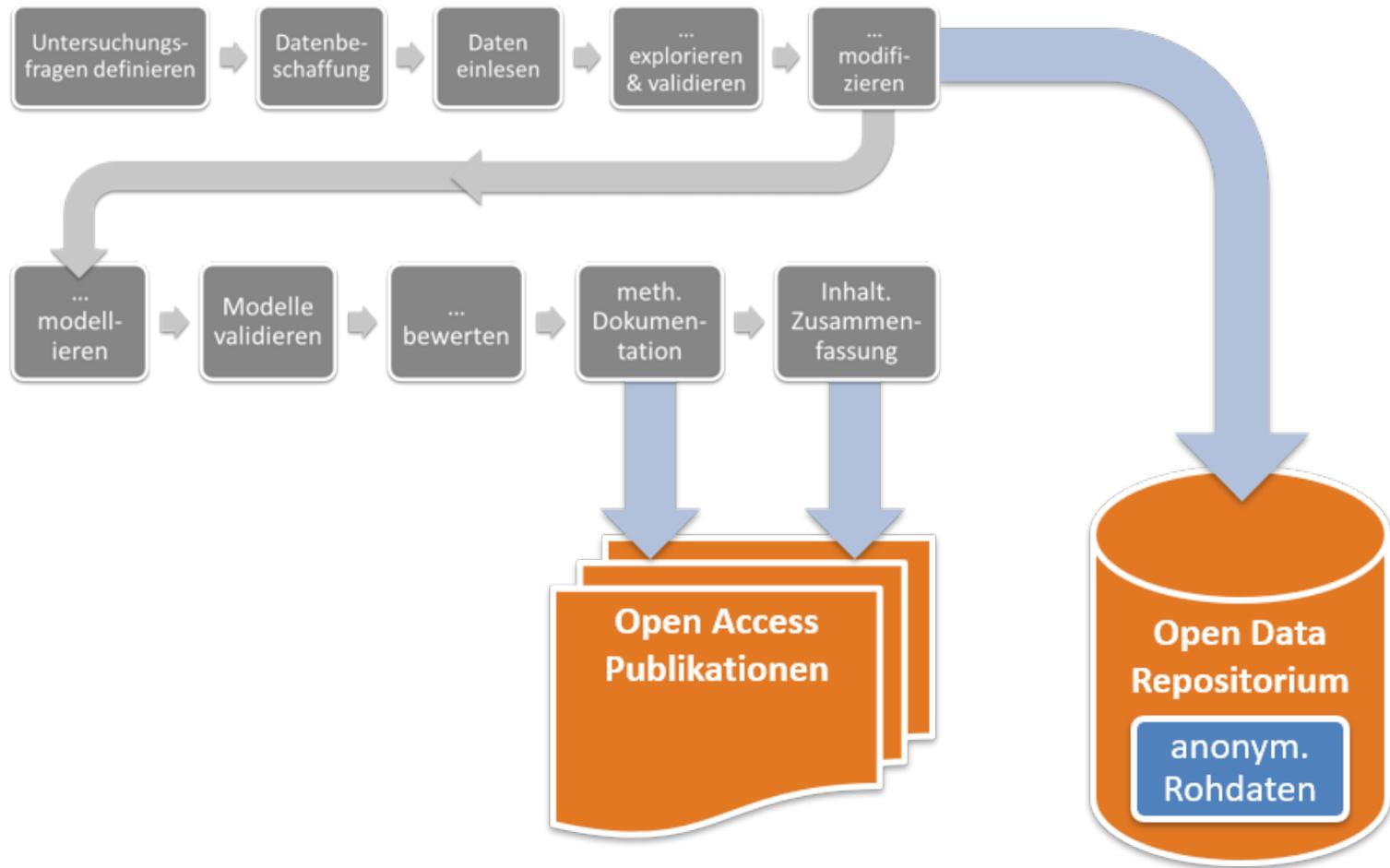
Aggregation



@ Klassifikation, Aggregation & Generalisierung



Ein idealtypischer (vgl. Open Science) Ablauf





Daten sind Daten, oder?

3. Versuch: Wooooooooooooo!



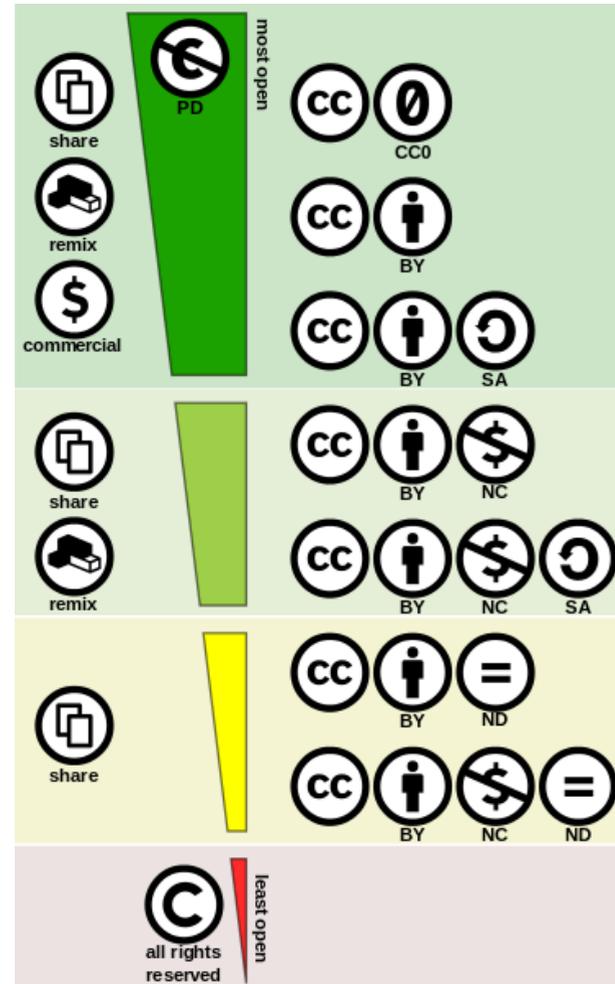
(Giphy)

Tirols Berge – ein Beispiel



Qualität & Eignung von Sekundärdaten

- Relevanz → Fragestellung
- Institutioneller Ursprung
- Nachvollziehbarkeit:
 - Erhebungsmethode
 - Objektivität, Reliabilität & Validität
 - Aufbereitung der Rohdaten
 - inhaltl. Verständlichkeit
- Nutzbarkeit → Lizenz



Strategien der Datenvalidierung

- Datentyp-bezogene Überprüfungen:
 - Konsistenz
 - Sonderzeichen etc.
 - Dezimalzeichen etc.
- Überprüfung von Spannweiten, Summen etc.
 - BSP. bei %
- Cross-Reference Check der Validität
 - Was sagen andere Datenquellen?
- Überprüfung der Konsistenz
 - Lücken etc.
- Überprüfung von Verweisen
- ...

Bedeutung von Open Data

"sämtliche Datenbestände, die im Interesse der Allgemeinheit [...] ohne jedwede Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden."

(Lucke & Geiger 2010:4)

- Zugang
- Weiterverbreitung
- Folgenutzung
- Abwesenheit technologischer Restriktionen (SIC!)
- Keine Einschränkung der Einsatzzwecke

Ein kleines Experiment:

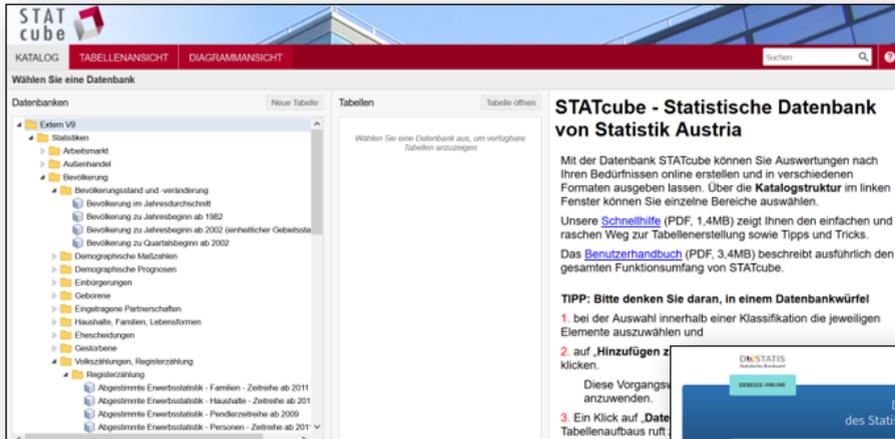
Tirol - Idyll in den Alpen ...

- Wie viele Gemeinden gibt es in Tirol?
- Wie viele Männer und Frauen leben in jeder dieser Tiroler Gemeinden?

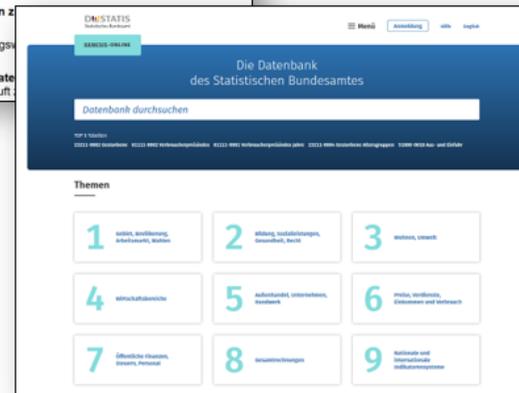


(barnyz, Flickr, CC BY-NC-ND 2.0)

Woher Sekundärdaten beziehen?



The screenshot shows the STATcube web application interface. At the top, there are navigation tabs for 'KATALOG', 'TABELLENANSICHT', and 'DIAGRAMMANSICHT'. Below the navigation, there is a search bar and a prompt to 'Wählen Sie eine Datenbank'. The main content area is divided into two sections: 'Datenbanken' on the left, which displays a hierarchical tree view of data categories such as 'Städtebau', 'Arbeitsmarkt', 'Außenhandel', 'Bevölkerung', and 'Erwerbstätigkeit'; and 'Tabellen' on the right, which contains a search bar and a list of available tables. A sidebar on the right provides introductory text about the database and links to a 'Schnellhilfe' (PDF) and a 'Benutzerhandbuch' (PDF).



The screenshot shows the Destatis website interface. At the top, there is a navigation bar with 'Destatis' and 'Statistik' links. Below the navigation, there is a search bar with the text 'Datenbank durchsuchen'. The main content area is titled 'Themen' and features a grid of nine numbered categories, each with a small icon and text: 1. Arbeit, Bevölkerung, Arbeitsmarkt, Wohnen; 2. Bildung, Beschäftigung, Gesundheit, Recht; 3. Medien, Umwelt; 4. Wirtschaftswachstum; 5. Arbeitsmarkt, Unternehmensstruktur; 6. Preise, Verdienste, Einkommen und Verbrauch; 7. Öffentliche Finanzen, Steuern, Sozialleistungen; 8. Gesundheitsleistungen; 9. Nationale und internationale Wettbewerbspositionen.

- <https://statcube.at>
- <https://www-genesis.destatis.de>

Woher Sekundärdaten beziehen?



GitHub

gesis
Leibniz-Institut
für Sozialwissenschaften

A screenshot of the Zenodo website. The top navigation bar is blue with the Zenodo logo, a search bar, and links for "Upload" and "Communities". A user profile dropdown shows "karl-michael.hoeferl@uibk.ac.at". The main content area is titled "COVID-19 related communities" and features a card for "Chicago COVID-19 Response" with a globe icon, "Browse" and "New upload" buttons, and a description. Below this is a section for "Featured uploads related to COVID-19" with a list of items, including "OpenAIRE Covid-19 publications, datasets, software and projects metadata" and "New York Times Coronavirus (Covid-19) Data in the United States".