

# LV Datengestützte Analysemethoden

## 3. DATEN NÜTZEN

Sommersemester 2018  
FH Joanneum Graz  
Studiengang Journalismus und Public Relations

Lehrender: Stefan Kasberger

**Stefan Kasberger**  
**@stefankasberger**



Dieses Werk ist lizenziert unter einer **Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International Lizenz**.

# **DATENANALYSE WORKFLOW**

# Workflow

- 1) Fragestellung und Anforderungen definieren
- 2) Daten sammeln
- 3) Daten bereinigen
- 4) Daten prozessieren
- 5) Daten analysieren
- 6) Ergebnisse interpretieren und dokumentieren

# 1. DESIGN

# 1. Design

**Was will ich mit dem Ergebnis machen bzw. durch die Analyse besser verstehen?**

→ z. B. Visualisierung, Entscheidung, Publizieren, Produkt verkaufen?

Schritte:

- Recherchieren, recherchieren, recherchieren: relevante Personen, Publikationen, Artikel, Tools, Daten, etc.
- Risikoabschätzung machen
- Randbedingungen und Sonderfälle mit bedenken
- Input von anderen einholen und die Gedanken öffnen

# 1. Design

## Beachte:

- nicht zu viele Daten und Analysen, sondern die Richtigen
- jeder Fehler beim Design kostet hinten hinaus das 10-100 fache an Zeit, Geld und Energie
- Faktoren Zeit und Raum

# 1. Design

## Outcome:

- Fragestellung und Anforderungen/Grenzen
- Weg zur Beantwortung → Methoden und Daten

## **Beispiel:**

Die Entwicklung der Hotel-Nächtigungen in Graz von 2005-2009 im Vergleich zu 2000-2004 von BesucherInnen aus dem EU-Ausland.

# DATEN SAMMELN



## **2a. DATEN ERSTELLEN**

# 2a. Eigene Daten erstellen

## Datenerhebungen

- Umfragen
- Interviews
- Beobachtung
- Recherche
- Messungen: Sensoren

# 2a. Eigene Daten erstellen

## Beachten

- Bias
- Repräsentativ
- Nutzung klären
- Sample Size
- neutrale Formulierungen
- Konfiguration Messinstrumente
- Metadaten

# **BEISPIEL METADATEN WETTERSTATION**

# Übung Metadaten

Stellt euch vor:

Ihr macht morgen eine einmalige Umfrage auf deutsch in der Herrengasse, wo ihr FussgängerInnen (w+m) nach deren Zeitungs-Lesegewohnheiten befragt. Ihr habt dazu einen Fragebogen mit, seit von 10-12 Uhr im Namen der FH unterwegs und speichert die Daten dann am Abend in eine CSV Datei. Weiters habt ihr von den Befragten das Einverständnis eingeholt, die Daten unter der Creative Commons Lizenz BY offen und anonymisiert frei zugänglich zu machen.

→ Erstelle die dazu passenden Metadaten als Key-Value Paare in einer metadata.json Datei. (5min)

## **2b. DATEN FINDEN**

# Wege

**1: Internet**

2: Bibliotheken: Sammelbänder,  
Enzyklopedien, Lexikas

3: Fach-Communities:

Konferenzen, Repositories,  
Meetups, Events, ExpertInnen

# Data Repositories

- [data.gv.at](https://data.gv.at) (offen)
- [opendataportal.at](https://opendataportal.at) (offen)
- Eurostat
- Statistik Austria
- World Bank (offen)
- Gapminder
- EU



**BEISPIEL ???**

# Übung Datenrecherche

## **Selber ein Dataset recherchieren:**

1) Suche die Ergebnisse der US-Wahl 2016.

2) Sieh dir die Daten an.

3) Sieh dir die Metadaten an

→ versuche es selber ohne Hilfe von anderen.

# Übung Datenrecherche

## **Stell dir folgende Fragen:**

- 1) Beinhalten die Daten alle benötigten Informationen um das Ergebnis zu verstehen?
- 2) Darfst du die Daten verwenden?
- 3) Reichen die Metadaten aus um die Daten zu verstehen?
- 4) Welche zusätzlichen Metadaten wären hilfreich gewesen?

**DATEN INSPIZIEREN**

# 1. Nutzung

Bevor die Daten verwendet werden dürfen, muss sicher gegangen werden, ob die Daten für den gewünschten Zweck verwendet werden dürfen.

- Lizenzierung erlaubt freie Nutzung
- Sondervereinbarung mit UrheberIn

## 2. Metadaten ansehen

Um den Inhalt des Datensets verstehen zu können, ist zumeist das Inspizieren der Metadaten notwendig.

# 3. Datei öffnen

Um die Daten zu verstehen und deren Qualität kontrollieren zu können, muss die Datei zuerst einmal mit der passenden Anwendung geöffnet werden.

# 4. Struktur

## **Welche Struktur haben die Daten?**

- Sind sie unstrukturiert/semi-strukturiert oder strukturiert?
- Handelt es sich um eine Tabelle, ein Netzwerk oder um Text?

→ mit Auge und Maschine



# 5. Vollständigkeit

**Ist das Dataset komplett?**

- Fehlt eine Beobachtung komplett?
- Fehlen einzelne Datenpunkte?

→ mit Auge und Maschine

# 6. Datentypen

**Welche Datentypen sind enthalten?**

- Skalierung
- IT
- Dimension

→ mit Auge und Maschine

# 7. Korrektheit

## **Sind die Daten fehlerfrei?**

- Anomalien suchen
- falsche bzw. unerlaubte Werte
- falsche bzw. unerlaubte Bezeichnungen

→ mit Auge und Maschine

# **3. DATEN BEREINIGEN**

# Bereinigen

## **„Unsaubere“ Daten bereinigen**

- Fehler in Strings
- Encoding
- Fehlende Werte (NaN)
- sehr zeitaufwändig.

# **5. DATEN PROZESSIEREN**

# Prozessieren

## Transformierung für die Analyse

- Datenstruktur verändern (z. B. Tabelle zu Dictionary)
- Berechnungen durchführen (Summe, Durchschnitt, etc.)
- mit anderen Daten anreichern
- Daten filtern, gruppieren, selektieren
- Daten exportieren

## Kontakt

[www.offenewahlen.at](http://www.offenewahlen.at)

[@stefankasberger](https://twitter.com/stefankasberger)

[stefan.kasberger@okfn.at](mailto:stefan.kasberger@okfn.at)

[www.okfn.at](http://www.okfn.at)

## UrheberInnenrecht:

Dieses Werk ist, sofern nicht explizit anders angegeben, lizenziert unter einer Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

Urheber: Stefan Kasberger (2018).

## Markenrecht:

Alle in dieser Präsentation genannten Marken und Produktnamen sind eingetragene Marken-/Warenzeichen der jeweiligen Hersteller beziehungsweise Unternehmen.