



# Lokalklima

R-Austausch, 31.08.2021  
Corinna Grobe





## AWEL & Statistisches Amt



**Gian-Marco Alt**

AWEL

Abteilung Luft,  
Fachexperte



**Jörg Sintermann**

AWEL

Abteilung Luft,  
Fachexperte



**Corinna Grobe**

Statistisches Amt

Datashop,  
Konzeption &  
Programmierung

Darstellungen werden in definierten Intervallen aktualisiert und nachgeführt

**Automatisierte** Darstellungen von Messwerten\* und ausgesuchten Auswertungen zum Lokalklima im ZHweb in Form **interaktiver** Grafiken & Tabelle.

“

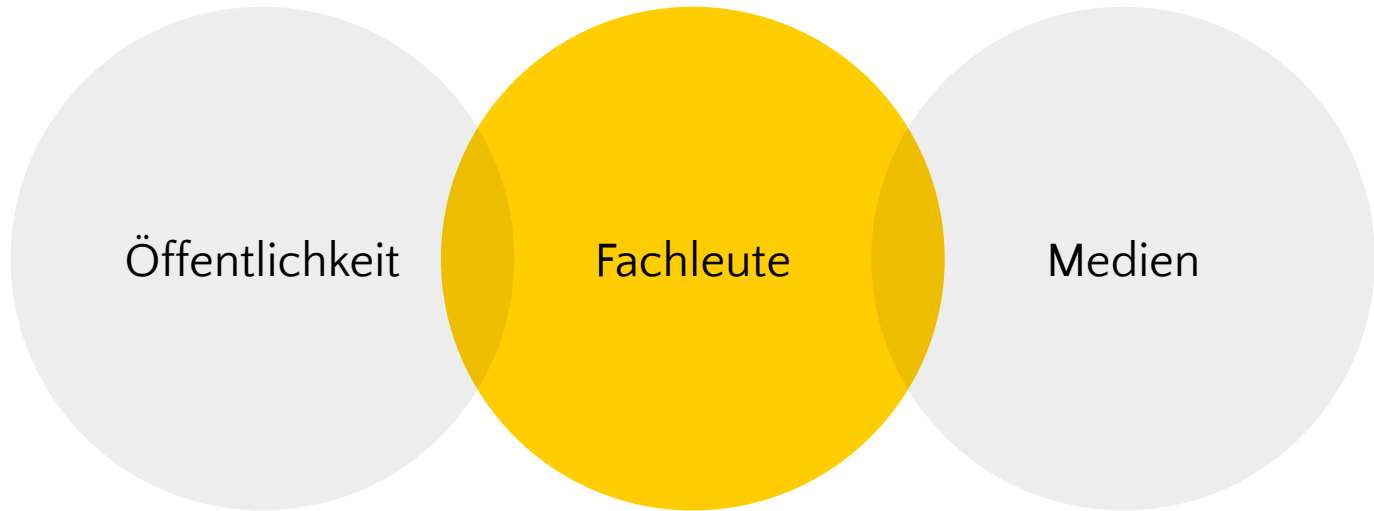
Benutzer können bestimmte Elemente der Anzeige selber auswählen (z.B. bestimmte Zeitperioden, Standorte, Standortkombinationen, Wärmeinseln etc.)

\* Das AWEL betreibt ein Messnetz von +/- 50 Sensoren im Kanton Zürich, welches die Temperaturwerte misst. Die Daten stehen **als offene Behördendaten** in einem offen und maschinen-lesbaren Format zur Verfügung.



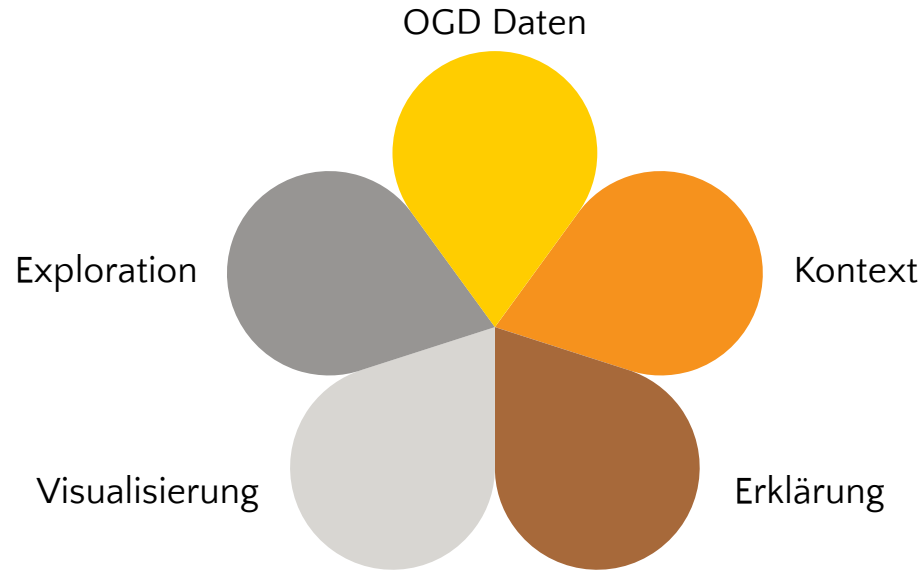
## Zielgruppe

---





## Bausteine



*Daten + Visualisierung + Narration*

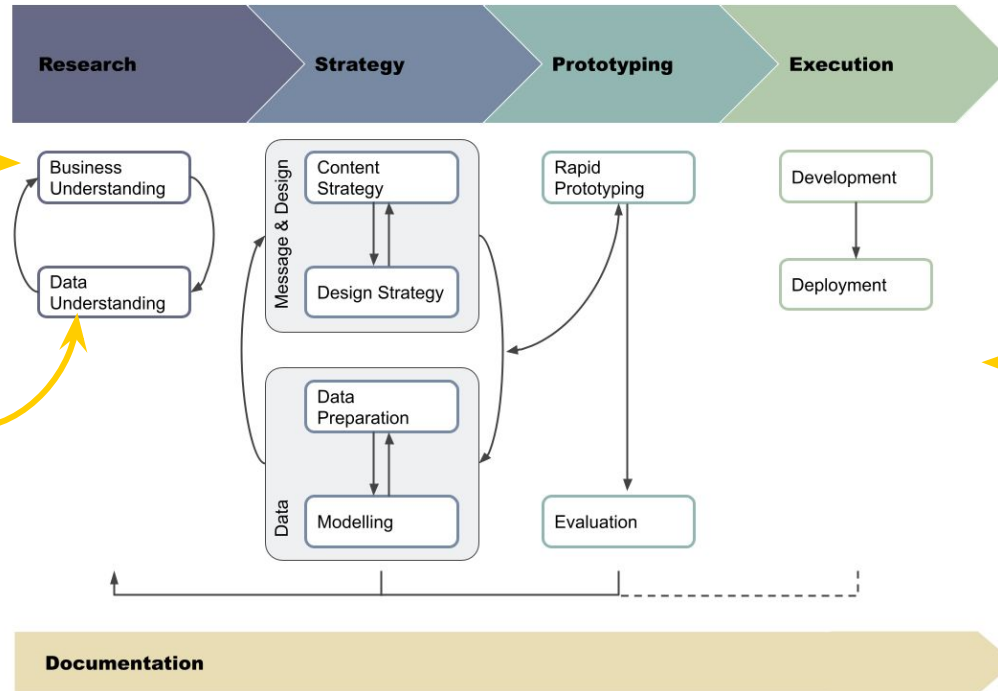




# Data Visualization Workflow

CRISP-DM\* &  
Design Process

Data Canvas



Shiny  
Dashboard von  
RStudio

Pen & Paper  
Sketches

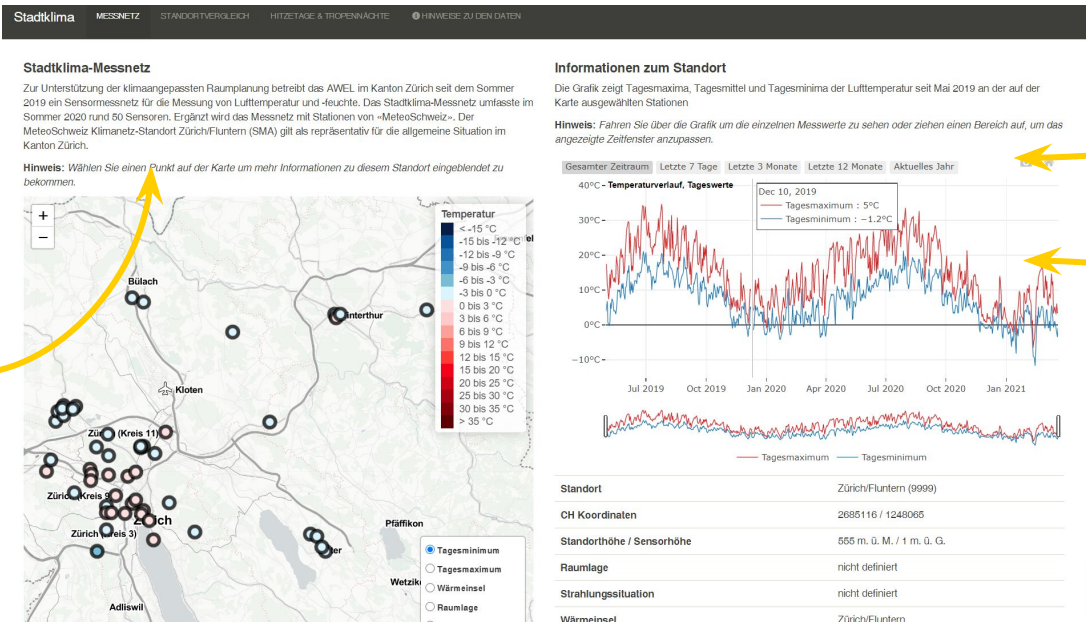
\* Cross Industry Standard Process for Data Mining



# Explain & Explore

Information  
strukturieren  
(Slideshow)

Kontext und  
Interpretation  
(Narration)



Nutzende führen  
(author-driven)

Richtiges Mass  
an Interaktivität  
(reader-driven)

Overview first - detail on demand





## R und Shiny Dashboard

- Skills: R wird bereits beim Kanton genutzt
- Support: R ist umfassend dokumentiert und aktive Community
- Wartung, Weiterentwicklung: Enabling der Data Owner
- Datenschutz: OGD, d.h. keine besonders schützenswerten Daten
- Hosting: Shiny Server – verfügbar, einfach, kostengünstig



# Data Cleaning

- Fehlende Werte → fortlaufende Zeitreihe
- Inkonsistenzen: Standortnamen und längere Kalibrationszeiträume → nicht-systematische Fehler

```
# DATA PROCESSING: POPULATE MISSING DATES -----  
## The problem is that we don't have all the dates during the tracked period in the data.  
## Instead, we have only the dates when values are actually measured.  
## To merge with the metadata we need all dates.  
  
## Reference: https://blog.exploratory.io/populating-missing-dates-with-complete-and-fill-functions-in-r-and-exploratory  
  
## To populate missing dates we are working with 'complete' from 'tidyr' package and 'seq.Date' from base R.  
## Basic structure: complete(Date = seq.Date(<start_date>, <end_date>, by=<date_unit>))  
tageswerte_populated <- tageswerte %>%  
  
  # Filling missing dates and values within each group of unique sensor / site combination  
  group_by(sensor, site) %>%  
  
  # 'seq.Date' works only for Date data type, so changing it by using as.Date  
  # mutate(date = as.Date(date)) %>%  
  
  # 'seq.Date' populates a sequence of Date data for the period that is configured  
  # 'complete' will add rows for the missing dates.  
  # Add all date column combinations that should be pre-filled  
  complete(date = seq.Date(min(date), max(date), by="day")) %>%  
  
  # Add all date column combinations that should be copied from the above line and filled into the newly added rows  
  fill(x, y, year, month) %>%  
  
  ungroup() %>%  
  
  # Joining WGS84 transformed coordinates  
  left_join(messnetz)  
  
## Check: We should have introduced NA for the calculated values  
na <- which(is.na(tageswerte_populated$T_min))  
  
check_new_rows <- tageswerte_populated %>% filter(is.na(T_min))
```



## Output Data

- Umrechnung der CH Koordinaten in WGS84
- Mergen der Datenquellen → möglichst standardisierten Datensatz
- Datenstruktur: Long vs. Wide Format
- Outputformat:

Data set	Output format	Size	Rows
OGD raw data	R object	319.9 MB	4,439,795 observations of 9 variables
	RDS	27.2 MB	
	CSV	381.8 MB	
The OGD raw data aggregated to hourly values	R object	66.1 MB	787,171 observations of 11 variables
	RDS	4.0 MB	
The OGD raw data aggregated to hourly values and narrowed down to the summer months	R object	18.3 MB	217,459 observations of 9 variables
	RDS	1.1 MB	
The OGD raw data aggregated to daily values for all combinations of sensor/site, filled in missing dates and added metadata as well the the MeteoSchweiz location	R object	4.5 MB	34,280 observations of 16 variables
	RDS	0.9 MB	



# Automatisierung

- Tägliche Aktualisierung
- R-Server: cronjob

```
1 setwd("/home/b105pcg@ji.ktzh.ch/git/AWEL_Lokalklima")
2
3 devtools::load_all(".")
4
5 source("scripts/1_data_import_ogd.R")
6
7 print(paste0("Letztes Update 1_data_import_ogd.R: ", Sys.time()))
8
9 source("scripts/2_data_import_meteoschweiz.R")
10
11 print(paste0("Letztes Update 2_data_import_meteoschweiz.R: ", Sys.time()))
12
13 source("scripts/3_transform_coordinates_ogd.R")
14
15 print(paste0("Letztes Update 3_transform_coordinates_ogd.R: ", Sys.time()))
16
17 source("scripts/5_merging_metadata_to_ogd.R")
18
19 print(paste0("Letztes Update 5_merging_metadata_to_ogd.R: ", Sys.time()))
20
21 source("scripts/7_calculate_t_diff.R")
22
23 print(paste0("Letztes Update 7_calculate_t_diff.R: ", Sys.time()))
24
25 source("scripts/8_data_export_zhweb_vis.R")
26
27 print(paste0("Letztes Update 8_data_export_zhweb_vis.R: ", Sys.time()))
28
29 files <- list.files("output")
30
31 file_paths <- paste0("output/", files)
32
33 purrr::walk(file_paths, ~upload_data(.))
34
35 print(paste0("Letztes abgeschlossenes Update: ", Sys.time()))
36
```



## Kontinuierlich verbessern

- Nativere Einbindung ins ZHWeb → Shiny App zu D3.js
- Reduzieren & Vereinfachen
- Weiterentwicklungen gemäss Backlog

A	B	C	D	E	F	G	H
ID	BESCHREIBUNG	BESCHREIBUNG ZIELZUSTAND	INSTANZ	TYP	STATUS	DATUM	ZUSTÄNDIG
1	Tab "Messnetz", X-Achse an Plot ist länger als der effektive Date Range	X-Achse geht von Min. bis Max. Datum	Applikation	Fehler	Umsetzung	2021-07-07	CG (STAT)
2	Anpassung Referenzstation	Falls die Daten der Meteoschweiz-Stationen tagesaktuell vorliegen, Anpassung Referenzstation auf Zürich/Affoltern	Applikation	Erweiterung	Offen	2021-07-12	GMA (AWEL)
3							
4							

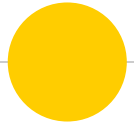


# Danke!

## Gibt es **Fragen?**

Ihr erreicht mich unter









- corinna.grobe@statistik.ji.zh.ch
- 043 259 75 07



# Appendix

Zusätzliche Slides zu den in der Präsentation gezeigten Inhalten

## Data Canvas

<b>Business Value</b>  <ul style="list-style-type: none"> <li>• Automatisch aktualisierte, interaktive Visualisierungen, Tabellen</li> <li>• Schnelle und offene Kommunikation zur Situation</li> </ul>	<b>Data Sources</b>  <ul style="list-style-type: none"> <li>• OGD: Messwerte, Metadaten</li> <li>• MeteoSchweiz: weitere Messstellen</li> </ul>	<b>Data Transformation</b>  <ul style="list-style-type: none"> <li>• Umrechnen der Koordinaten</li> <li>• Umgang mit fehlende Messwerten</li> <li>• ...</li> </ul>	<b>Obstacles</b>  <ul style="list-style-type: none"> <li>• Ausreichende Datenabdeckung</li> <li>• Datenqualität Metadaten</li> <li>• Datenmenge / Performance</li> </ul>	<b>Users</b>  <ul style="list-style-type: none"> <li>• Experten: Fachleute, AWEL</li> <li>• Nicht-Experten: Journalisten, Bevölkerung</li> </ul>
	<b>Metrics</b>  <ul style="list-style-type: none"> <li>• Tmin, Tmax, Tdiff, Tropennacht, Hitzetag etc.</li> <li>• Verteilungen je Wärmeinsel, Raumlage, Periode</li> </ul>	<b>Data Target</b>  <ul style="list-style-type: none"> <li>• Merge Tabelle mit Informationen aus allen Quellen</li> </ul>	<b>Frequency</b>  <ul style="list-style-type: none"> <li>• Sommer: wöchentlich</li> <li>• Restl. Jahr: monatlich</li> </ul>	



