



**CASUS**

CENTER FOR ADVANCED  
SYSTEMS UNDERSTANDING

# A scalable pipeline for effective forecast of COVID-19 in Germany, Czechia and Poland.

**Wildan Abdussalam**

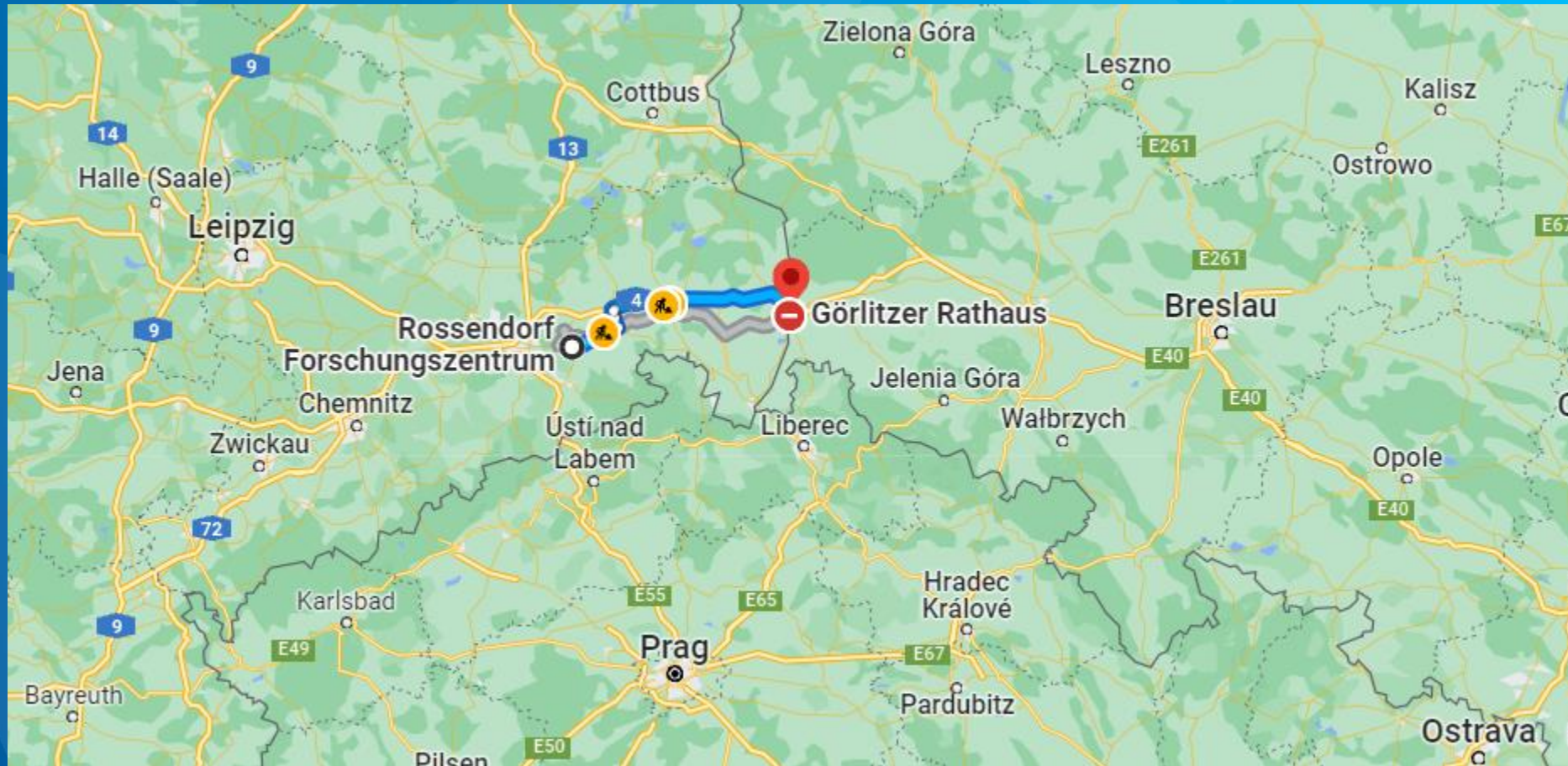
[www.casus.science](http://www.casus.science)

# Location





# Location



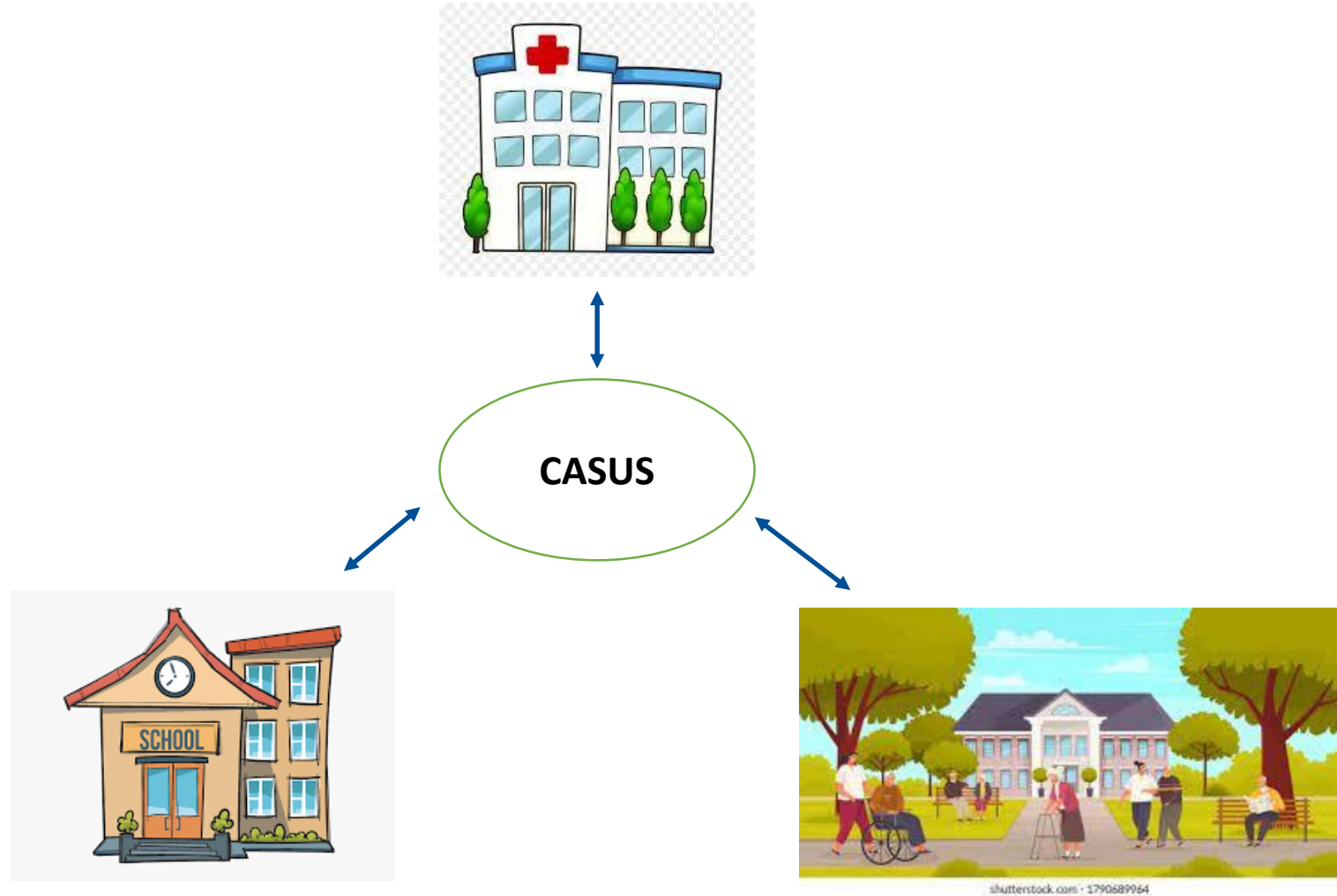


# Location



# Digital Health

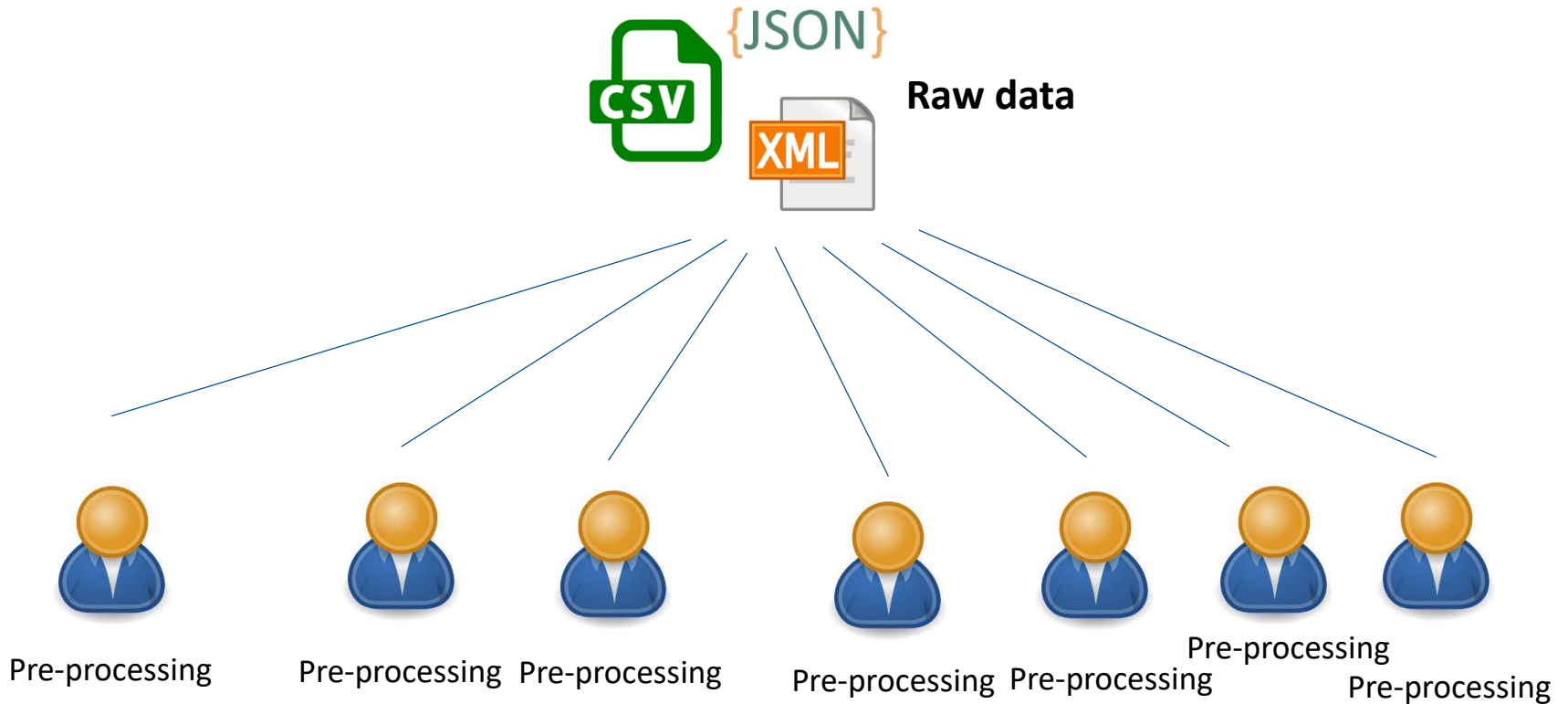
## Datasets Pre-requisites



# Data science activities

An example of analysing COVID-19 cases.

## Repository of Robert Koch Institute



Thanks To



# Pros cons of the previous slide

csv, xml and json

Pros	Cons
Used everywhere	Not inherently secure
Large user community	Susceptible to trivial human errors
Familiar User interface	Difficult to troubleshoot & test
Many built-in and 3rd party functions	Not designed for collaborative work
Easy learning curve	Trouble in handling large datasets
Independent work	Not built with Business Continuity in Mind
quick analysis of smaller datasets	Expensive in pre-processing datasets.

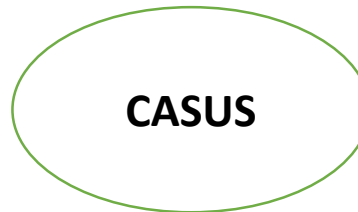




# Digital Health

## Datasets Pre-requisites

- ✓ Collaborative work
- ✓ Data security
- ✓ Large datasets
- ✓ Continuity
- ✓ Less data pre-processing for users
- ✓ Data Synchronization



### SQL:

- ✓ Preserve data integrity
- ✓ Process data quickly
- ✓ Store data securely
- ✓ Store large datasets
- ✓ Have an audit trail on the database



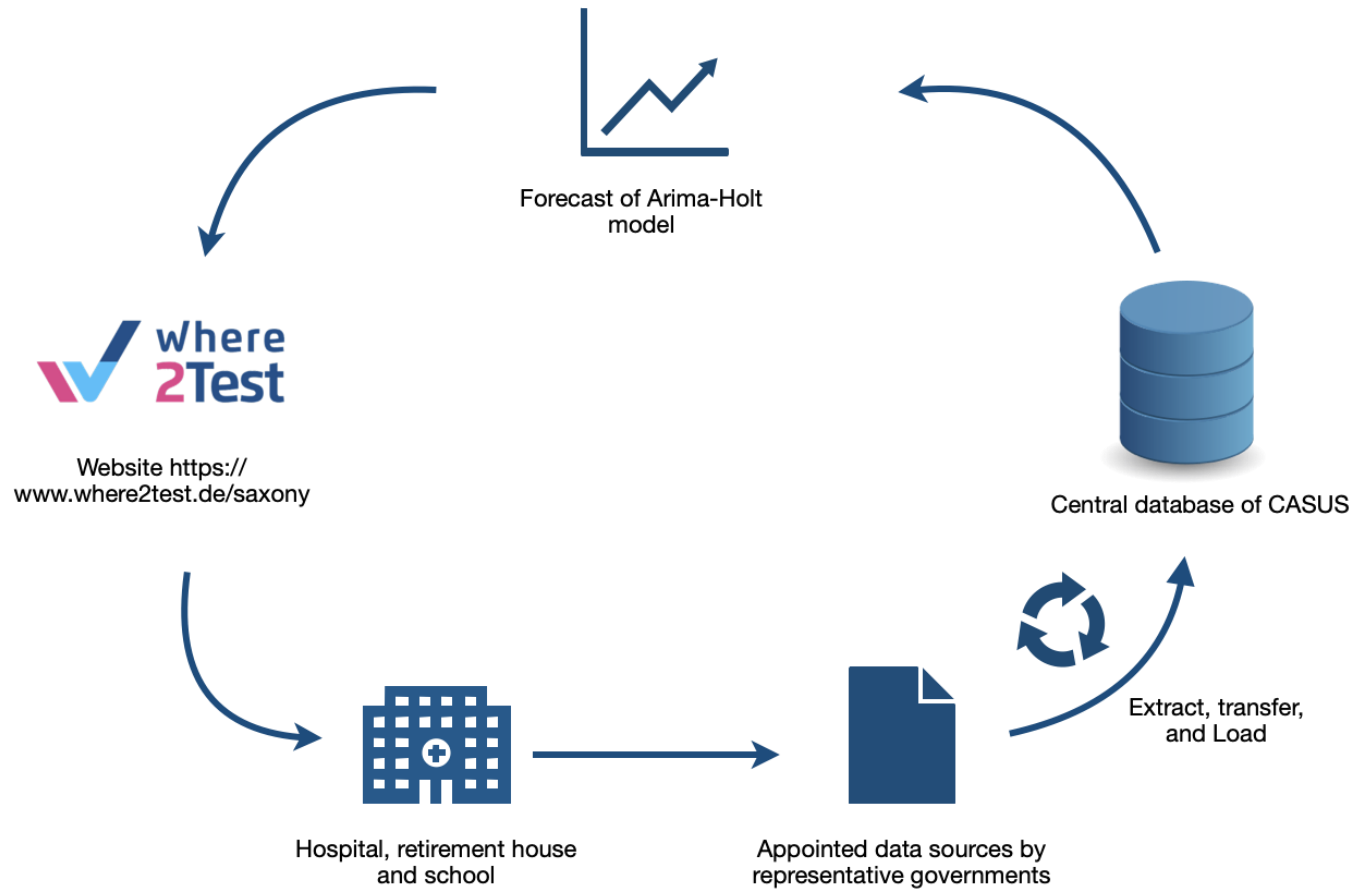
shutterstock.com · 1790689964





# Objective

## Data Integration Pipeline of COVID-19



# Challenges

## Data Integration of COVID-19 datasets

01

Various table  
structures of  
datasets

02

Various  
datafile  
formats

03

Data  
synchronizati  
on



# Efforts

## Data Integration of COVID-19 datasets

### EU

<https://qap.ecdc.europa.eu/public/extensions/covid-19/covid-19.html#eu-eea-daily-tab>

### Italy

<https://www.mdpi.com/1660-4601/17/15/5596>

### Maryland

<https://coronavirus.maryland.gov/>

### EU

<https://github.com/covid19-eu-zh/covid19-eu-data>

### UK

<https://coronavirus.data.gov.uk/details/interactive-map/cases>

### South America

[https://github.com/DataScienceResearchPeru/covid-19\\_latinoamerica](https://github.com/DataScienceResearchPeru/covid-19_latinoamerica)

### Germany

<https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/Landkreise/>

### John-hopkins

[https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)





# Background

## Data Integration of COVID-19 datasets

### Our concerns

- × Available dashboards are limited to continents, nations, states, and counties.
- × Lack of forecasting features.
- × Forecasting studies are limited to specific time windows.

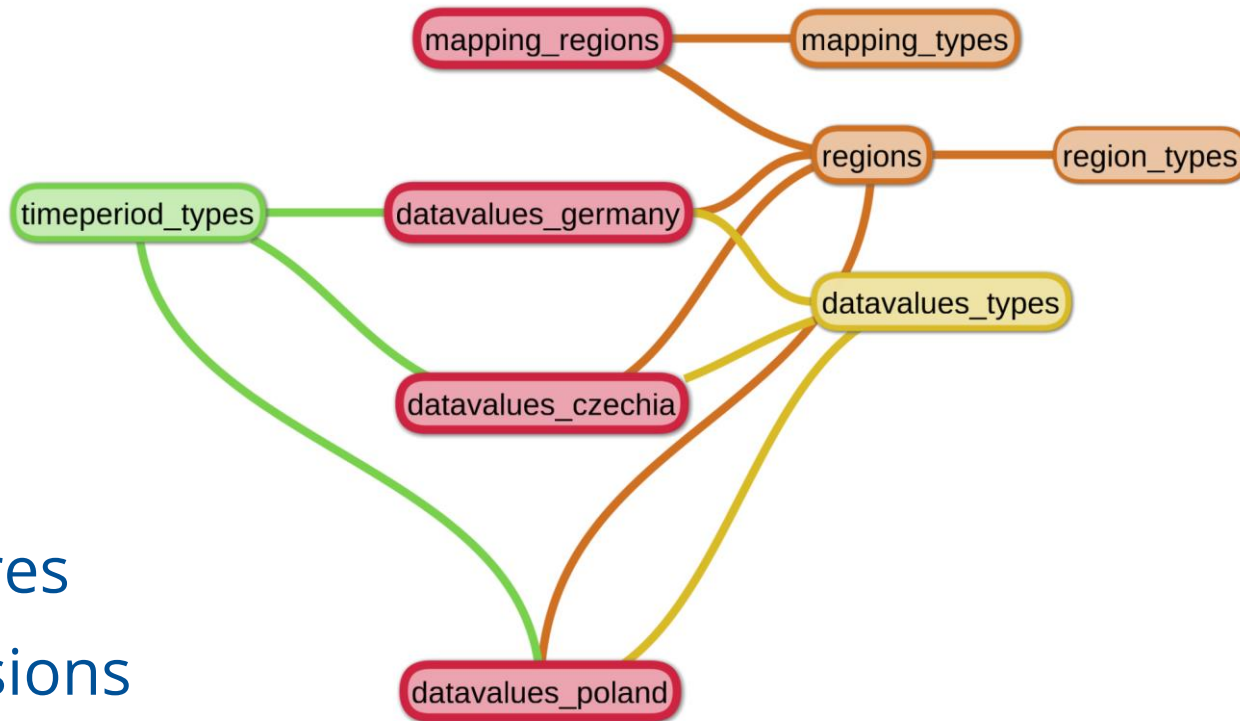
### Our interests

- Allow inter-country top-down spatiotemporal observation.
- Adding forecasting features.
- weekly updated forecasting based on daily data.



# Model

## Dimensional fact model



Facts

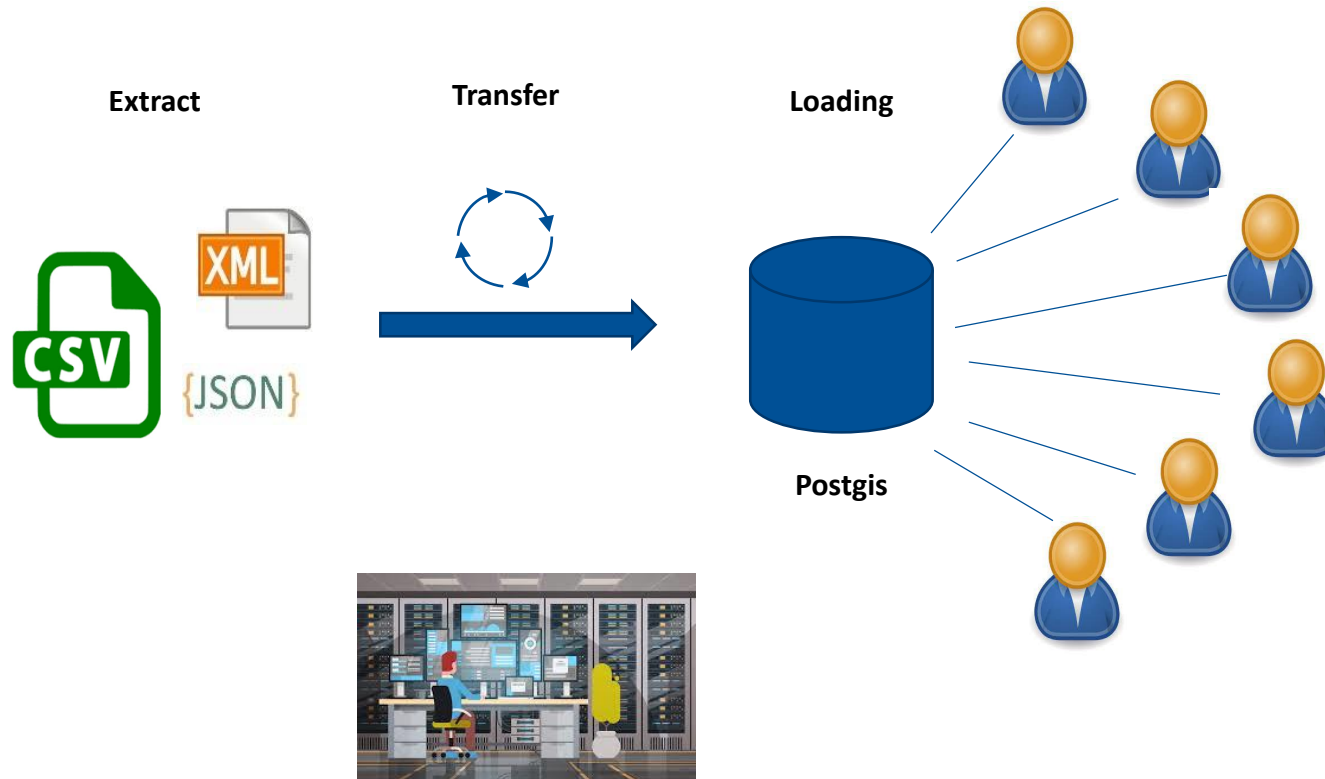
Measures

Dimensions



# Extract and transfer

## Transferring datasets from xml, csv and json to postgresql



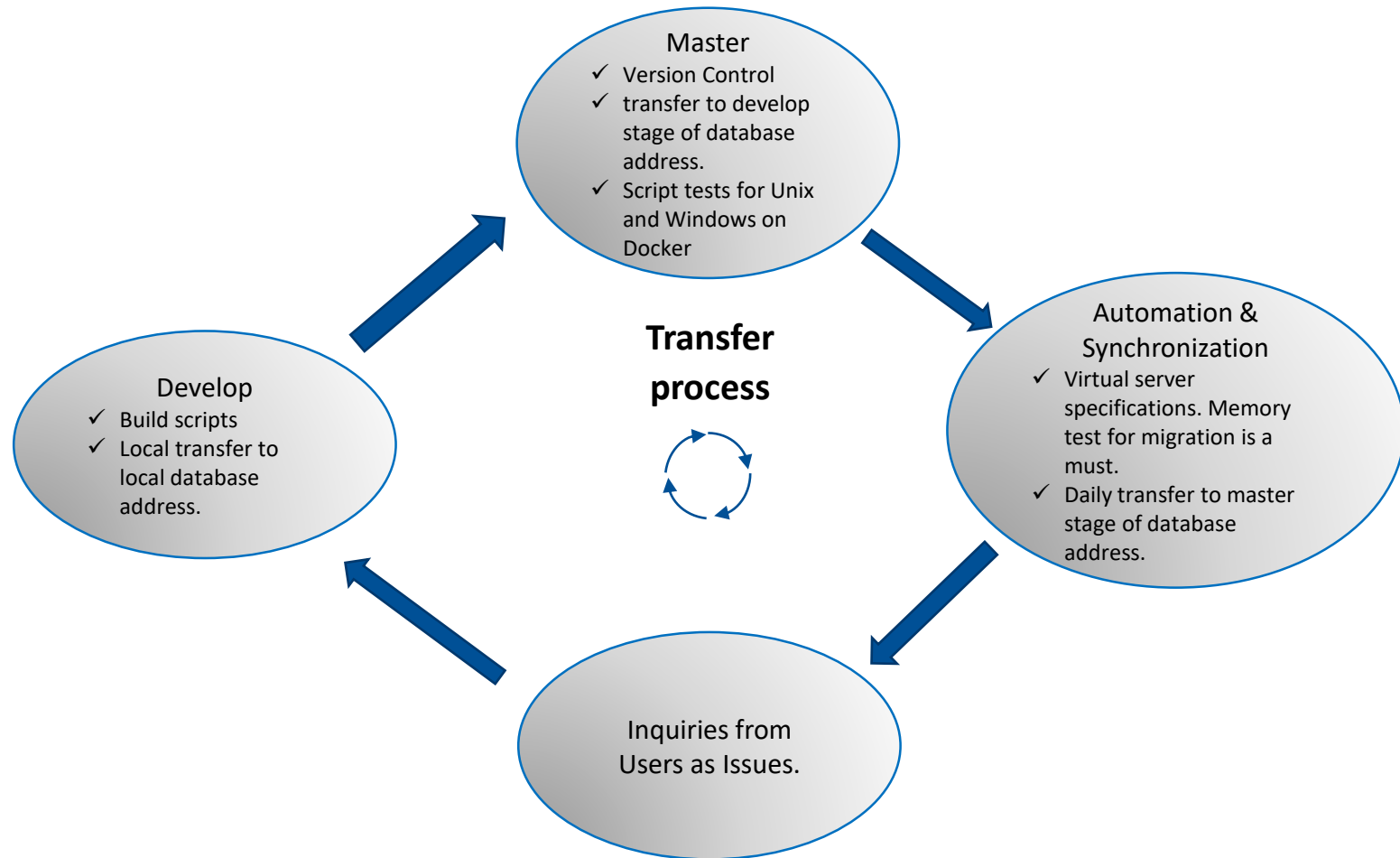
Pre-processing data  
by database admin.





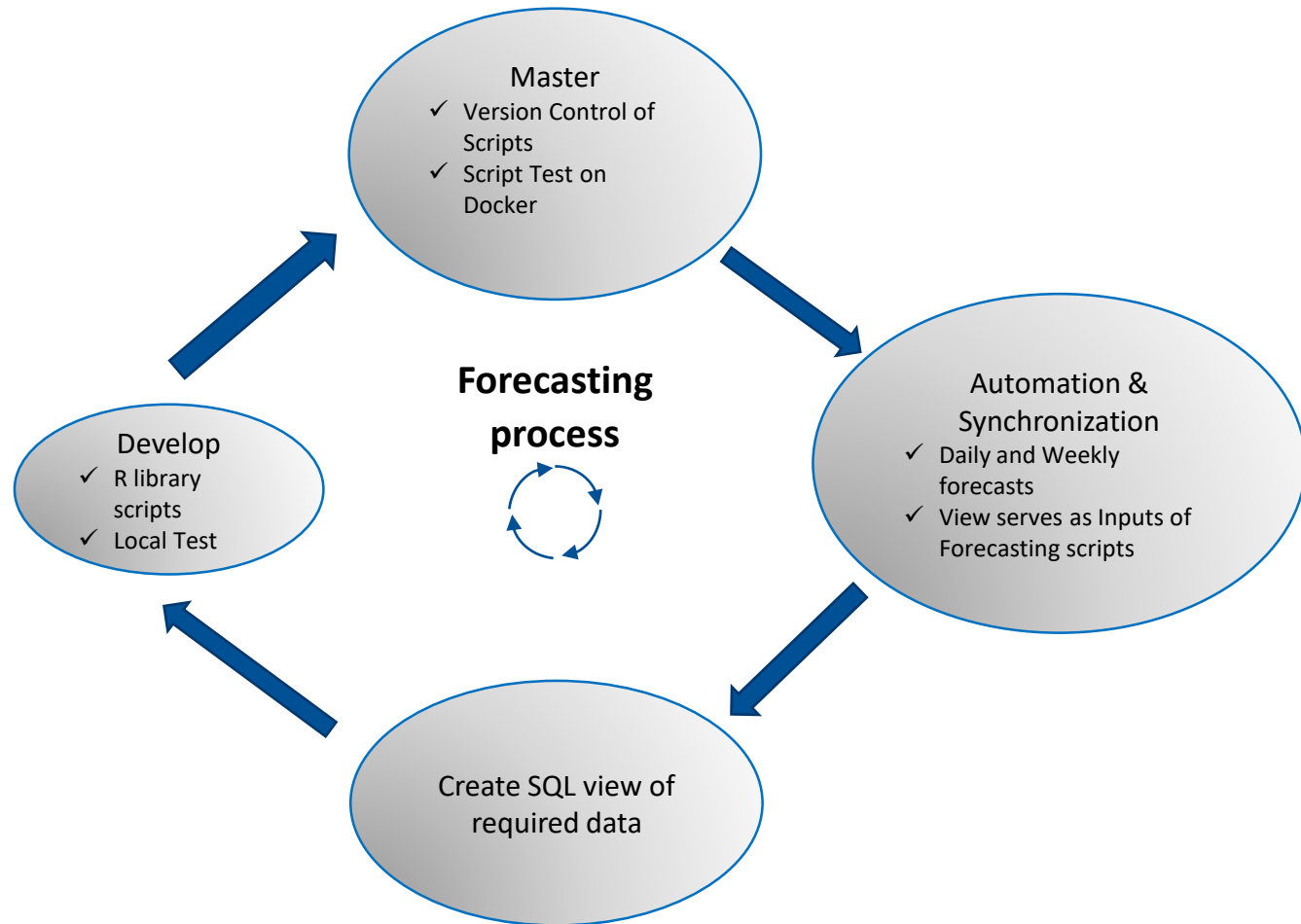
# Data transfer workflow

## Transfer process



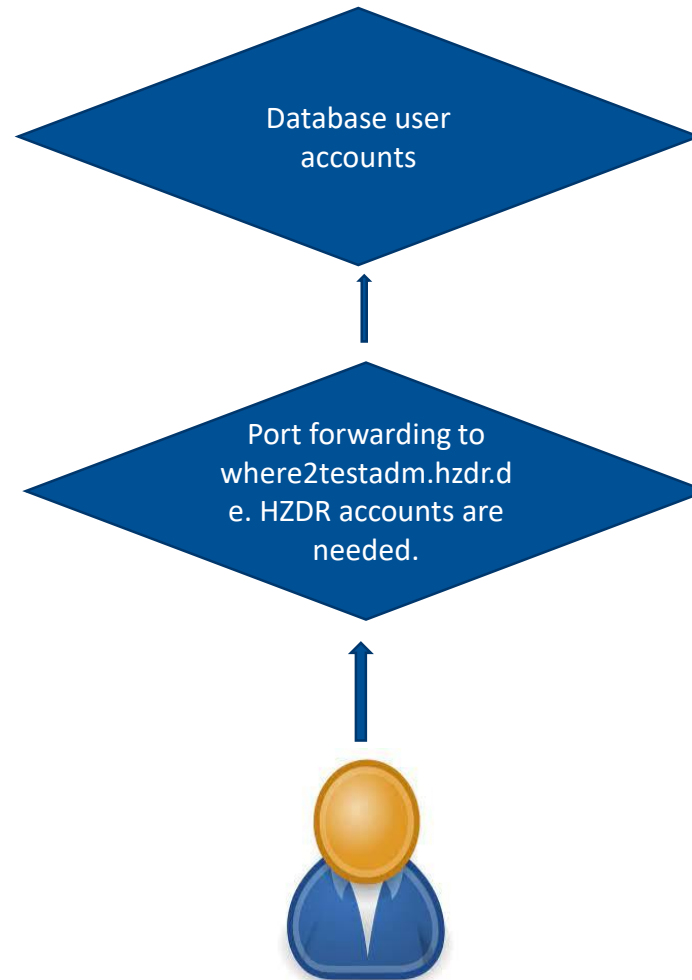
# Forecasting model workflow

## Arima Holt-Winters model



# Data security

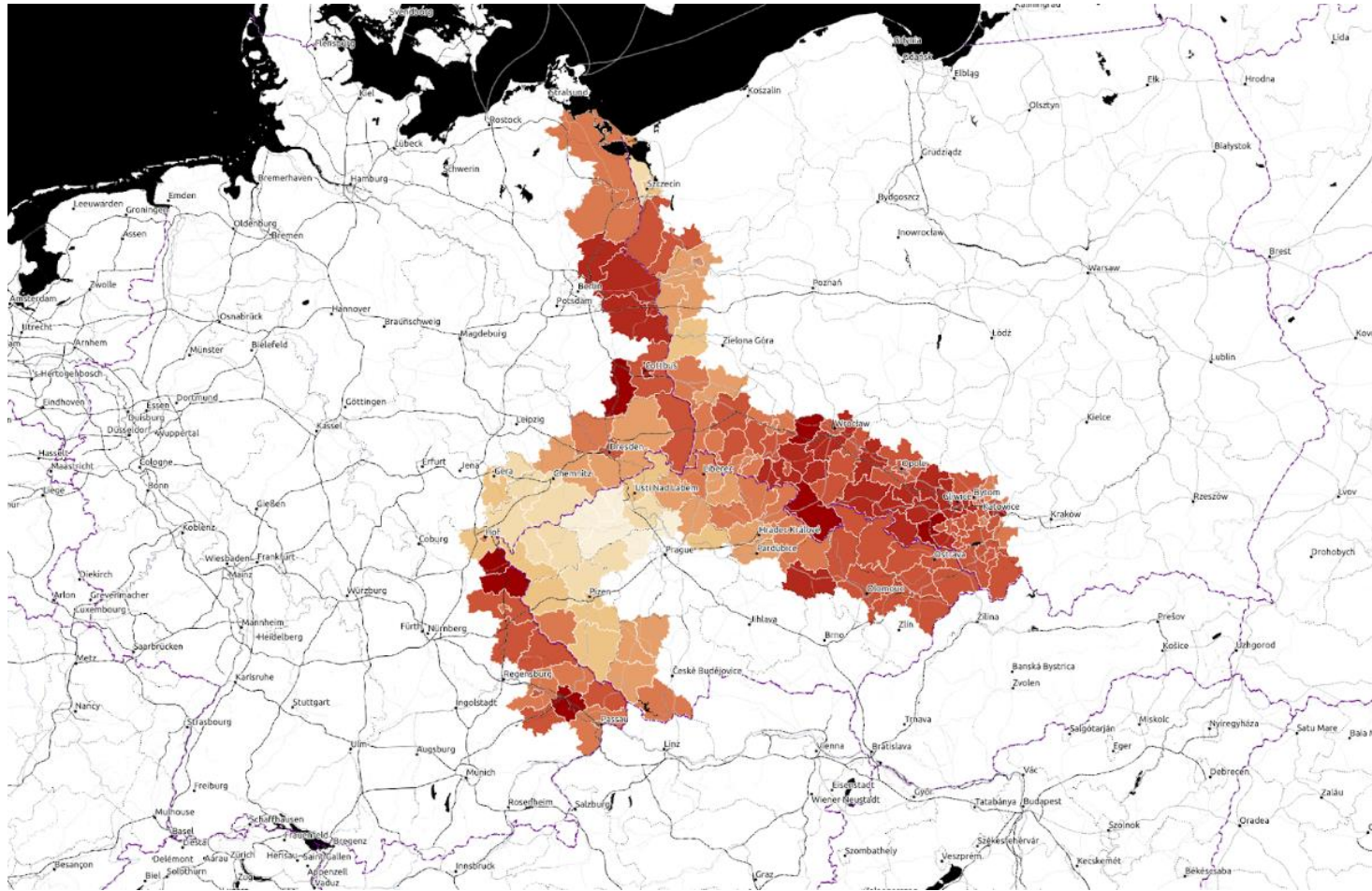
## User access





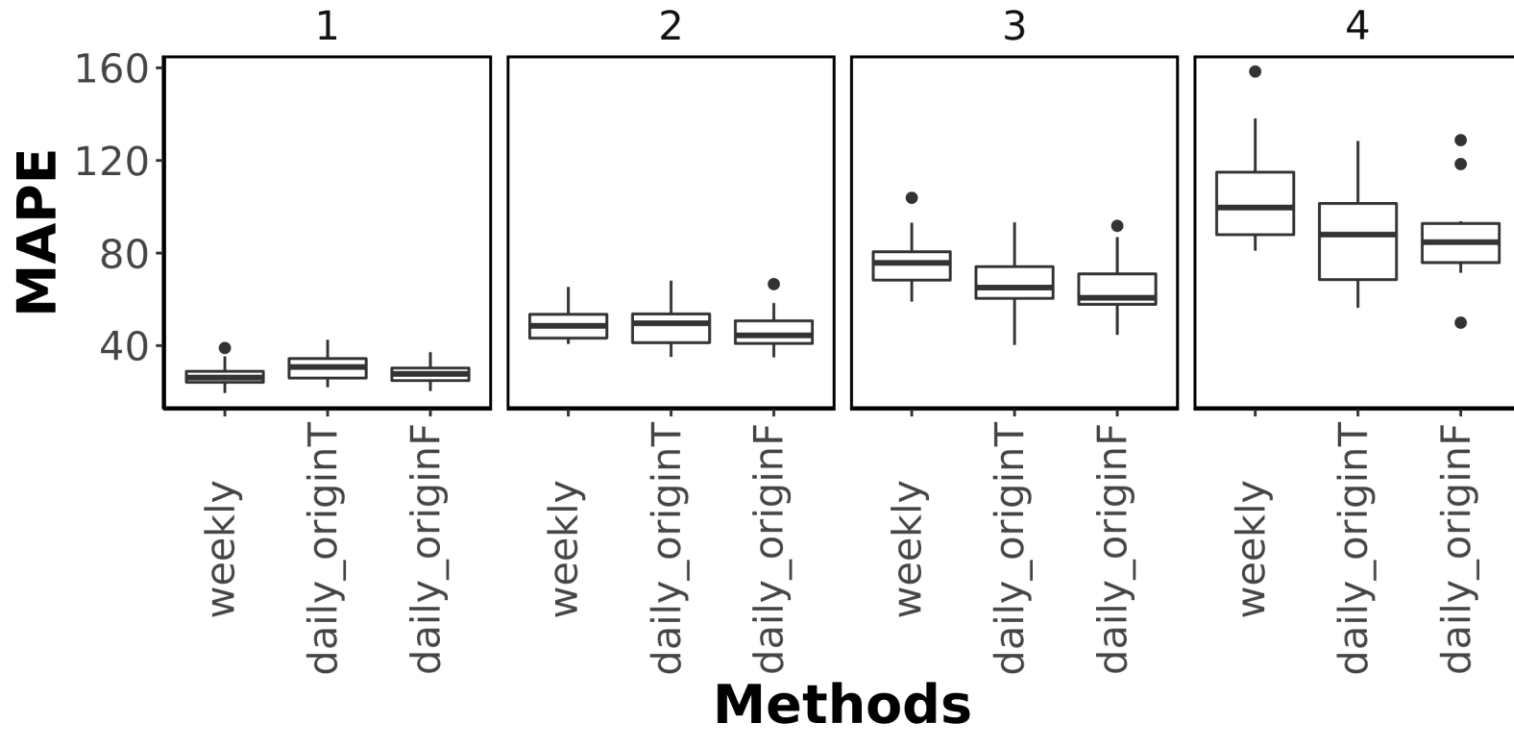
# Database usage for frontend

## Border effects for Germany, Czechia and Poland



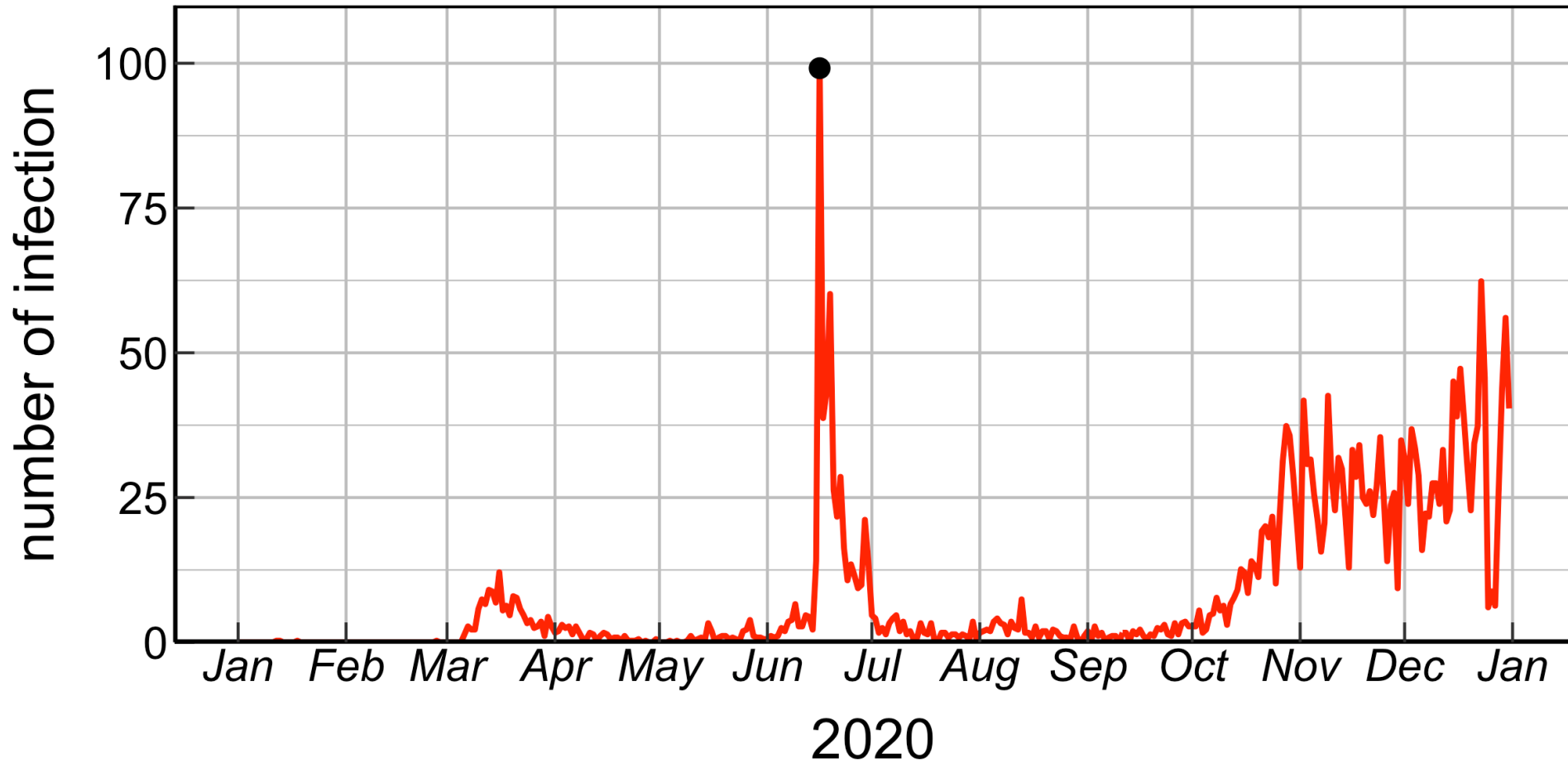
# Database usage for frontend

Forecast using (S)ARIMA-Holt model



# Database usage for frontend

## Outlier detection for superspreading events





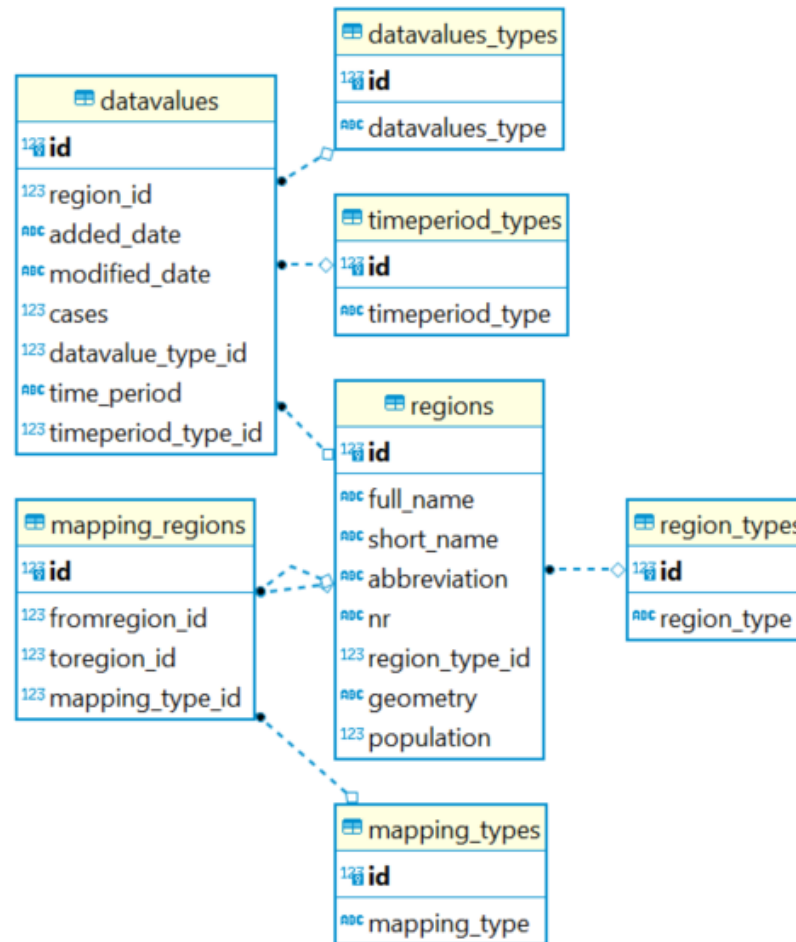
# CASUS

CENTER FOR ADVANCED  
SYSTEMS UNDERSTANDING

[www.casus.science](http://www.casus.science)

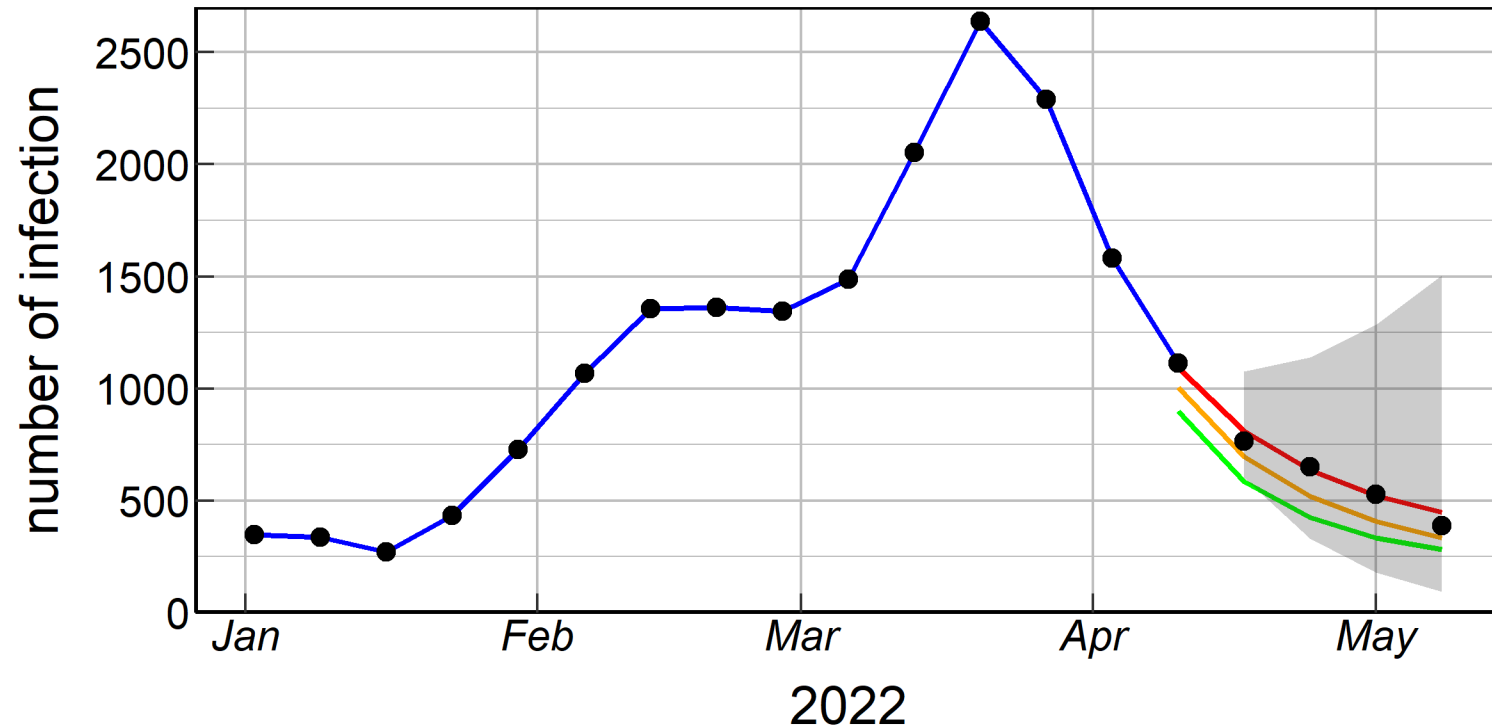
# Model

Synchronization of different table structures from data sources.



# Database usage for frontend

Forecast for Germany, Czechia and Poland





# Database usage for frontend

## Forecast using (S)ARIMA-Holt model

**Table 1** The Mean absolute percentage error (MAPE) of four horizons among various tests.

Model	Horizon	Weekly		Org (i)		Double (ii)		5week (iii)		Daily		weekly_now4week (v)	
		T	F	T	F	T	F	T	F	T	F	T	F
ARIMA	1	27	29	29	31	35	37	51	46	51	43	41	37
	2	53	54	60	59	61	63	99	80	100	78	74	69
	3	82	81	101	84	94	86	173	117	174	116	110	103
	4	112	109	158	108	138	108	286	154	286	153	145	136
	1 - 4	68	68	87	70	82	74	152	99	153	98	93	86
Holt	1	27	31	27	27	33	36	43	40	43	40	39	39
	2	52	59	52	51	55	61	78	69	79	70	71	69
	3	81	88	80	74	78	82	125	101	126	104	109	106
	4	112	117	113	98	104	103	180	136	180	140	150	146
	1 - 4	68	74	68	62	68	70	107	86	107	88	92	90
Mix	1	27	31	27	27	33	36	44	40	44	40	40	39
	2	52	58	52	52	56	61	80	69	81	70	71	69
	3	81	87	80	74	78	82	128	101	130	103	109	105
	4	112	116	113	97	104	102	187	135	188	138	149	145
	1 - 4	68	73	68	63	68	70	110	86	111	87	92	90