Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Data Management for Epidemiology

Borchert, Dr. Schapranow

Data Management for Digital Health
Winter 2023

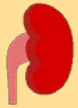
Agenda
Pillars of the Lecture

Hasso Plattner Institut

Medical Use Cases


Biology Recap


Oncology


Nephrology

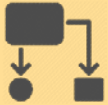

Infectious Diseases

Technology Foundation


Data Sources


Data Formats


Processing and Analysis


Software Architectures

Machine Learning

Data

ML

Refine

Evaluate

Prediction + Probability

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

Agenda
Pillars of the Lecture

Medical Use Cases


Biology Recap


Oncology


Nephrology


Infectious Diseases

Technology Foundation

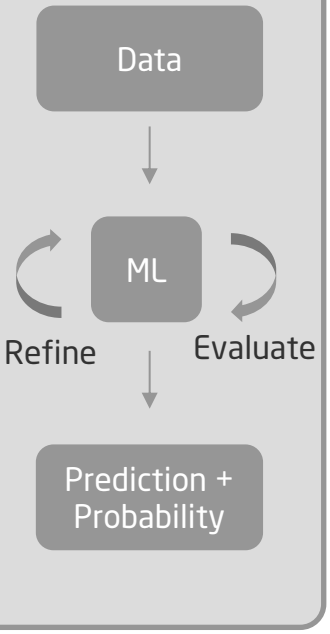

Data Sources


Data Formats

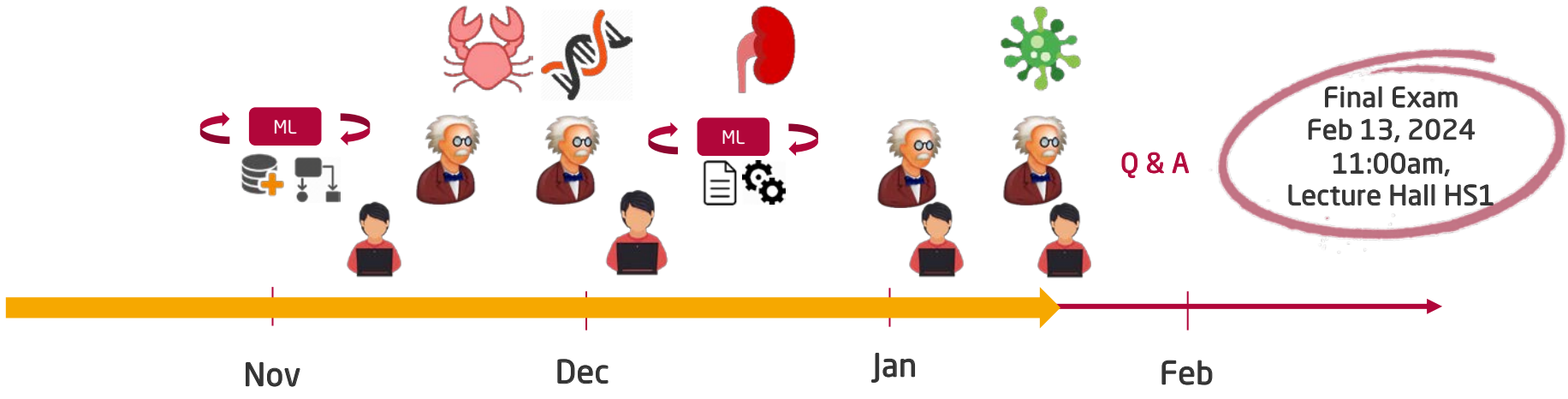

Processing and Analysis


Software Architectures

Machine Learning



Lecture Schedule



- Lecture Kickoff
- Actors in Healthcare
- Digital Health Data

- Machine Learning (ML) Foundations
- Use Case Oncology
- Biology Recap

- Natural Language Processing
- Use Case Nephrology & Intensive Care
- Supervised ML & Deep Learning

- Use Case Infectious Diseases
- Unsupervised ML

Agenda

- Epidemiology and Disease Modeling
- Contact Tracing
- Epidemiological Surveillance and Situational Reports
- Variant Surveillance

Epidemiology
Definitions You Should Know

- Epidemiology := Study of risk factors, occurrences, distributions, causes and consequences of diseases in a defined population
- Makes use of statistical methods / models to describe spread of diseases and identify cause-effect relationships

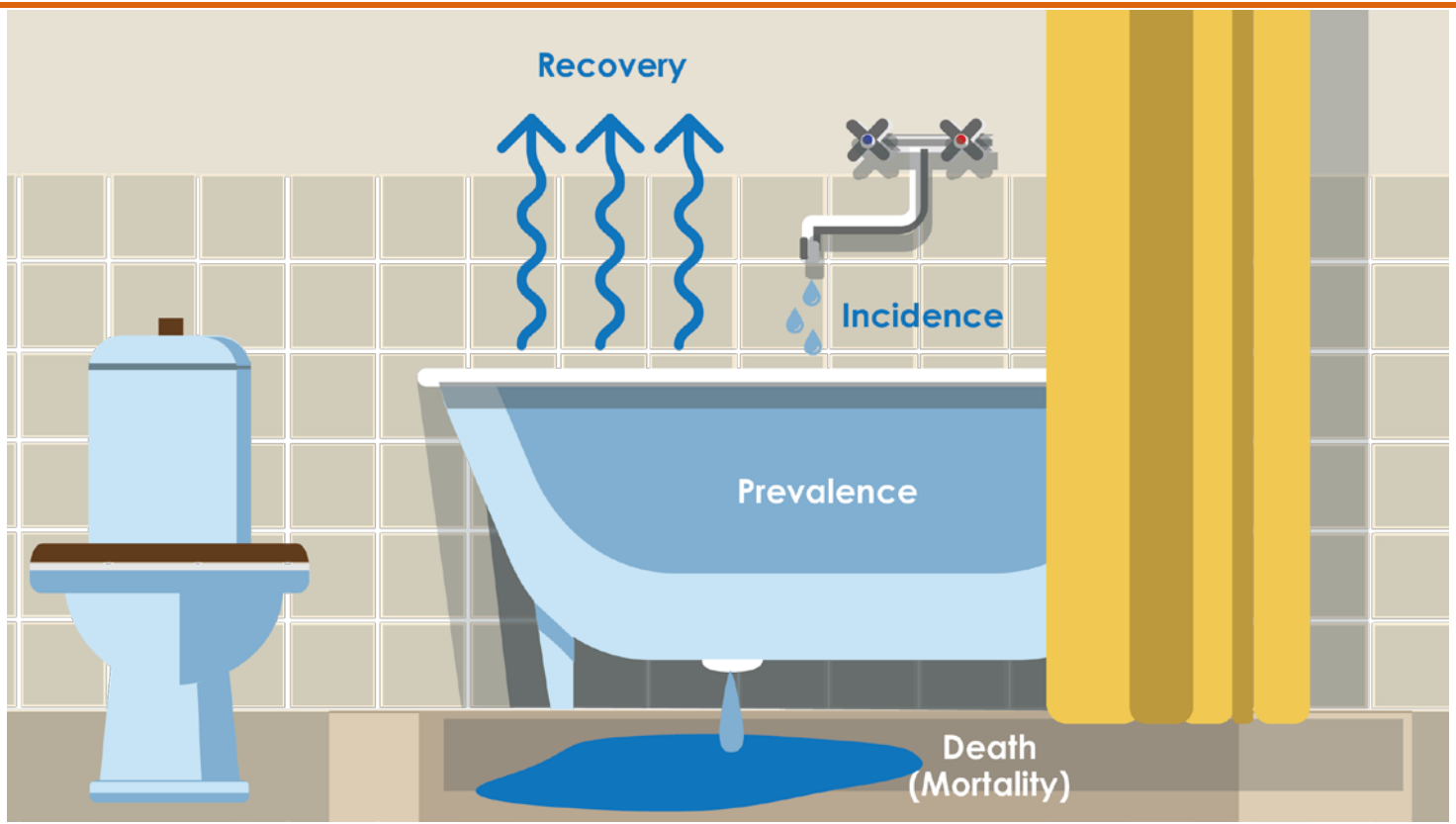
Prevalence and Incidence
For a specific medical condition in a given population…

- (Period, Point) Prevalence: $Pr = \frac{n_{diseased}}{n_{pop}} \rightarrow [0,1]$
  - Ratio of all cases and the complete population for a given timespan or at a given point while event of interest have happened before
- Risk, Cumulative Incidence, Incidence Proportion: $IP_{\Delta t} = \frac{n_{diseased}\, during\, \Delta t}{n_{risk}\, at\, t_0} \rightarrow [0,1]$
  - "Probability to catch the disease"
  - Ratio of new cases during a given timespan Δt and persons at risk at t_0 (excl. infected person at t_0).
- Incidence Rate: $IR = \frac{n_{diseased} during\, \Delta t}{\sum t \cdot nrisk} \rightarrow [0, \infty[$ (per normalized pop) per timespan
  - Ratio of new cases during a given timespan Δt and person years at risk during the same period of time, i.e. total amount of time (in years) each person of the population is at risk of the disease during the period of interest

Prevalence and Incidence Summary

	Prevalence	Incidence Proportion	Incidence Rate
Numerator	#all cases (new and pre-existing)	#new cases	#new cases
Denominator	Population at point in time / period of time	#population at risk	Total person years of observation

Prevalence and Incidence
Epidemiologist's Bathtub



https://www.technologynetworks.com/immunology/articles/incidence-vs-prevalence-329073

Example

- Scenario: As an epidemiologist, you conducted a survey amongst your fellow students. You ask whether people are currently infected with COVID-19.

Students / Years	2020	2021	2022	2023	2024
A	-	-	-	n/a	n/a
B	-	-	+	-	-
C	-	-	+	-	-
D	-	+	+	-	-
E	-	+	+	+	-

- Period prevalence for 2020-2024: ?
- Incidence proportion in 2021: ?
- Incidence proportion in 2022: ?
- Incidence rate: ?

Example (cont'd)

- Scenario: As an epidemiologist, you conducted a survey amongst your fellow students. You ask whether people are currently infected with COVID-19.

Students / Years	2020	2021	2022	2023	2024
A	-	-	-	n/a	n/a
B	-	-	+	-	-
C	-	-	+	-	-
D	-	+	+	-	-
E	-	+	+	+	-

- Period prevalence for 2020-2024: 4 total cases / 5 size of population = 80%
- Incidence proportion in 2021: 2 new case / 5 people at risk at t_0 = 40%
- Incidence proportion in 2022: 4 new case / 5 people at risk at t_0 = 80%
- Incidence rate: 7 new cases / 16 person years = 437,5 new cases per 1k person years

Compartment Models in Epidemiology

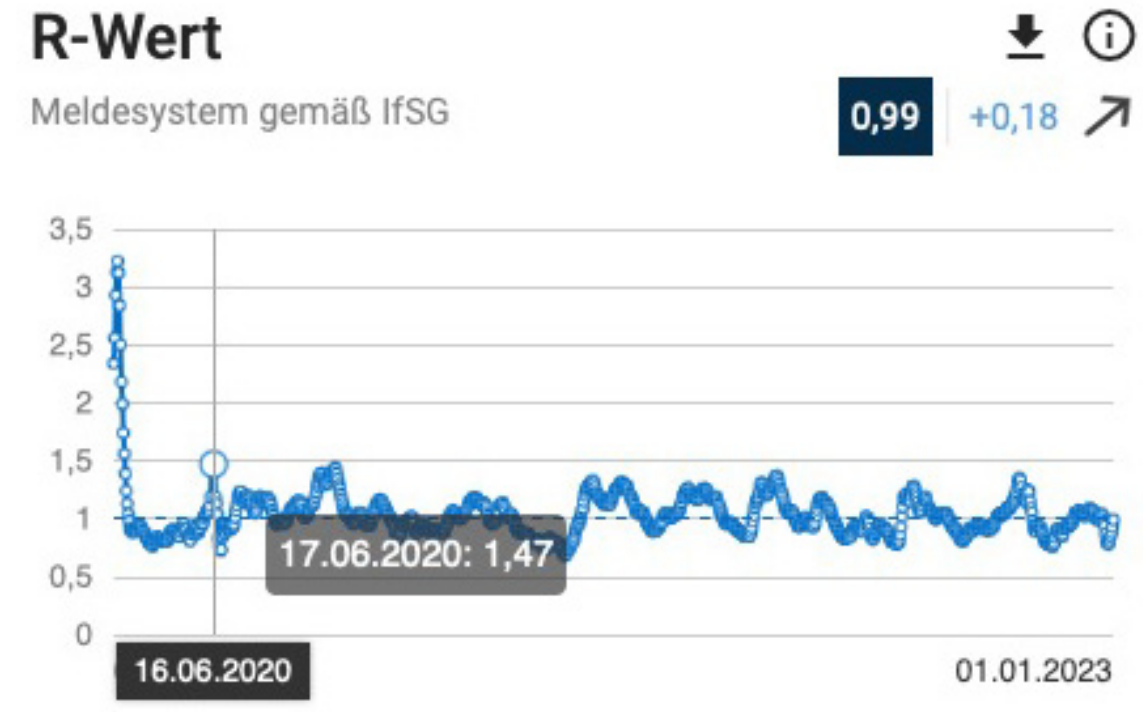
- Susceptible individuals
- (Exposed individuals)
- Infected individuals
- Recovered individuals
- S(E)IR model : = S → (E →) I → R

- Carrier individuals
- Deceased individuals
- Vaccinated individuals
- Maternally-derived immunity
- Further models: SECIR, SEIRD, MSIR, SIRV, etc.

Reproduction number R

- Basic reproduction R_0 (typically estimated) := Expected number of new cases caused by a single case at t_0 when all individuals were in compartment S
- Effective reproduction R_t (observed) := Avg. number of new cases caused by a single case at time point t (this is what you find in situational reports)
- Linking R_t and R_0: Let s be the proportion of people in compartment S, who can get infected (e.g. no immunity): $R_t = R_0 * s$
- Herd immunity := Indirect protection against infectious diseases once a specific percentage p_{immune} of the population has become immune so that $R_t < 1$.
  - $R_t < 1 \Leftrightarrow R_0 * s < 1$
  - $R_0 * (1-p_{immune}) < 1$
  - $p_{immune} > 1-R_0^{-1}$

Reproduction number R_t
COVID-19 in Germany Mar 2020 - Dec 2022

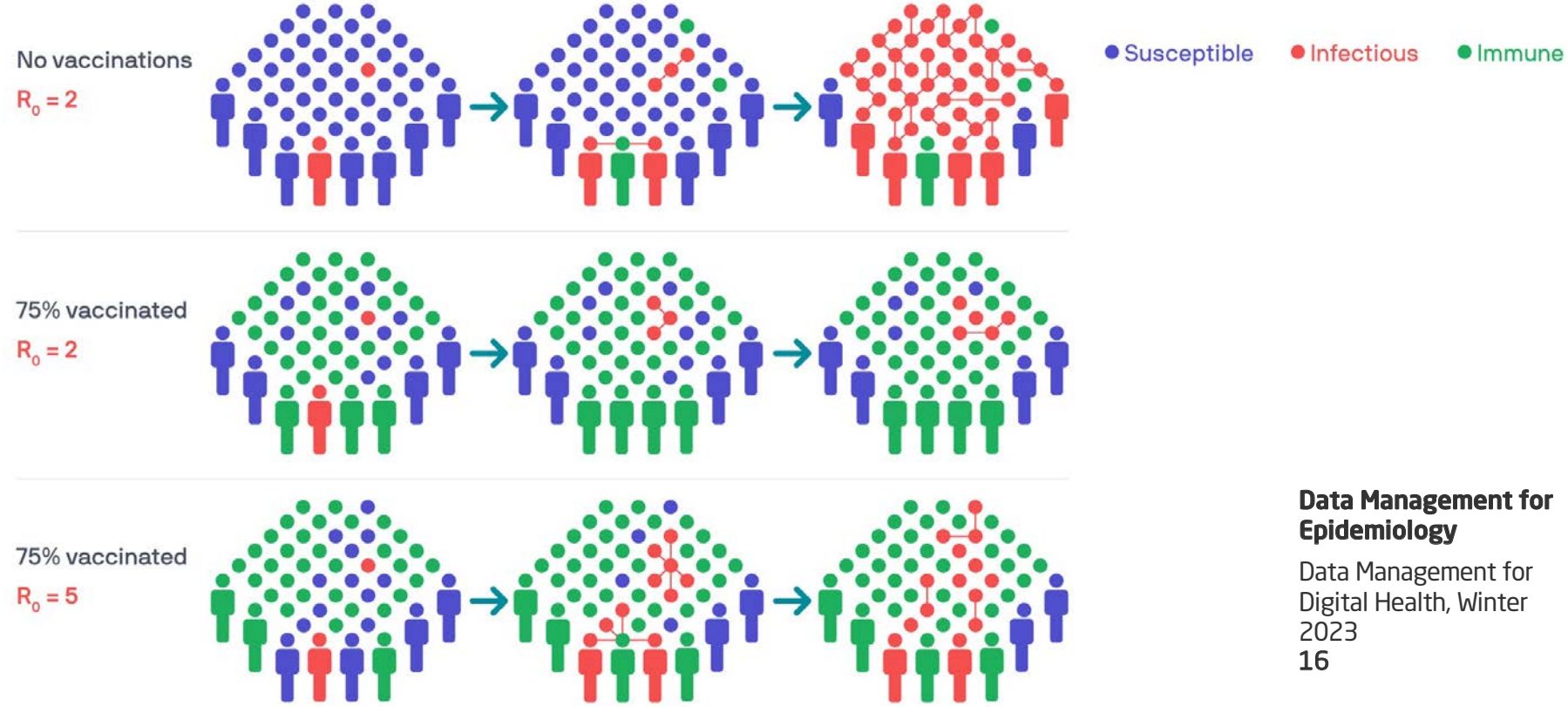


https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/COVID-19-Trends/COVID-19-Trends.html?__blob=publicationFile#/home

Reproduction number R
Examples

Disease	Transmission way	R_0	Herd immunity $(1-R_0^{-1})$
Chickenpox, mumps	Aerosol, Respiratory droplets	10-12	90-92%
AIDS	Body fluids	2-5	50-80%
COVID-19, SARS	Respiratory droplets, aerosols	2-4	50-75%
Ebola (2014)	Body fluids	1.8	44%
Influenza (seasonal strains)	Respiratory droplets	1.3	23%

Reproduction number R
Impact of p_{immune} / Vaccinations



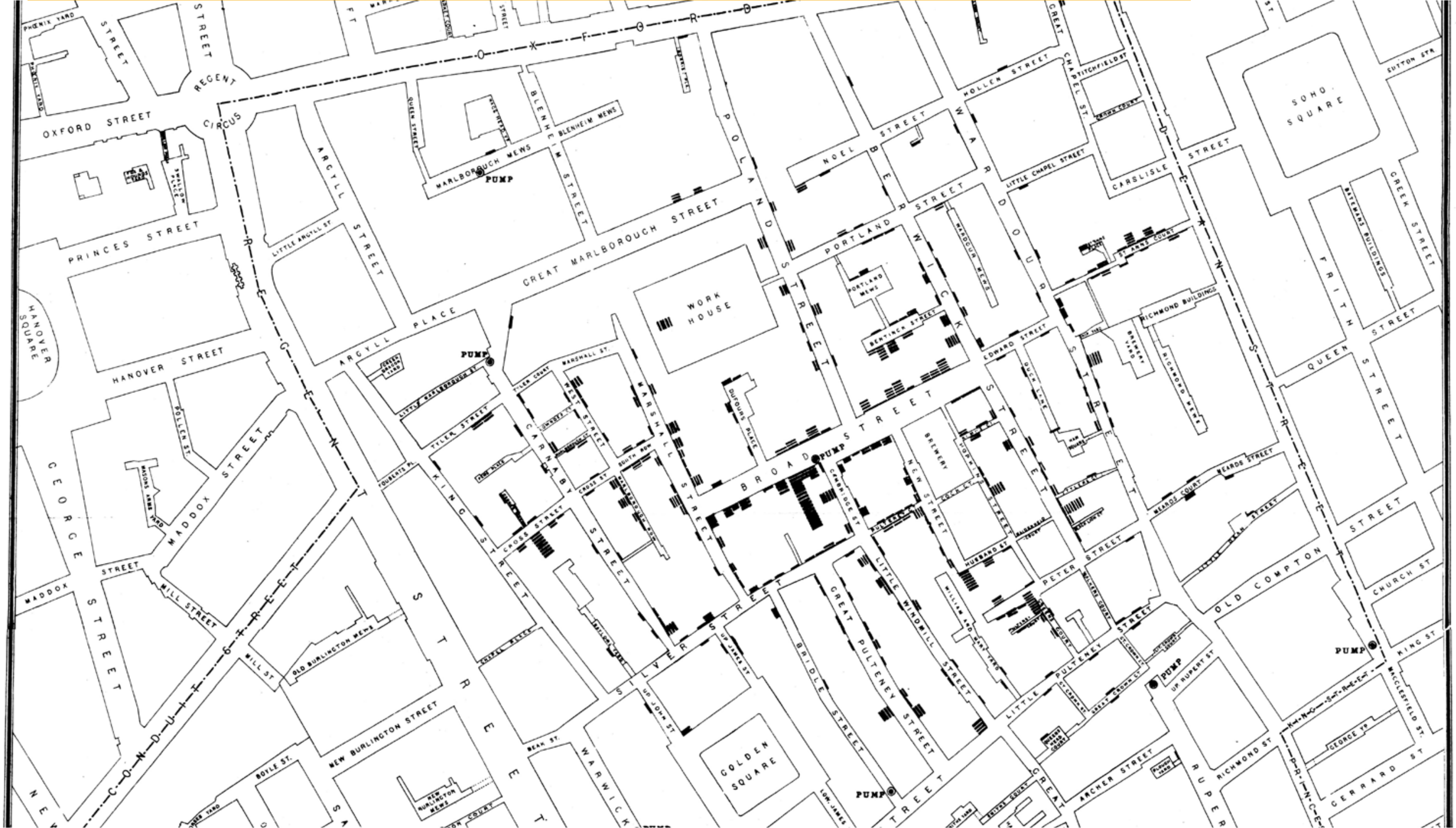
https://www.path.org/our-impact/articles/understanding-journey-herd-immunity/

Epidemiological Surveillance and Situational Reports
<< Brainstorming Sessions >>

- Epidemiological surveillance := On-going, systematic collection, analysis, evaluation and dissemination of health data for the purpose of planning, implementation and evaluation of disease control measures (RKI)

- What could be sources of health data?
- What could be measures?

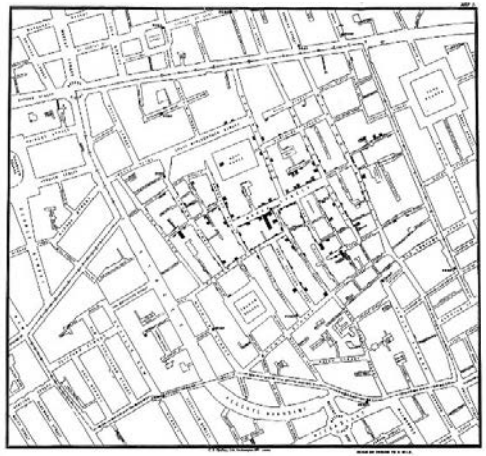
Photo by Jason Blackeye

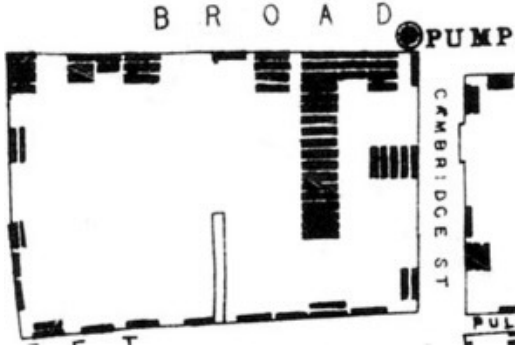


OXFORD STREET
REGENT CIRCUS
PRINCES STREET
HANOVER SQUARE
HANOVER STREET
GEORGE STREET
MADDOX STREET
MILL STREET
CONDUIT STREET
ARGYLL STREET
ARGYLL PLACE
LITTLE ARGYLL ST
GREAT MARLBOROUGH STREET
MARLBOROUGH MEWS
BLENHEIM MEWS
BLENHEIM STREET
PUMP
POLAND STREET
NOEL STREET
BERWICK STREET
WARDOUR STREET
PORTLAND STREET
PORTLAND MEWS
MOLLEN STREET
LITTLE CHAPEL STREET
CARSLISLE STREET
DEAN STREET
SOHO SQUARE
SUTTON STR
FRITH STREET
CREEK STREET
QUEEN STREET
RICHMOND BUILDINGS
RICHMOND MEWS
ST ANNS COURT
WORK HOUSE
BENTINCK STREET
EDWARD STREET
DUFOURS PLACE
BROAD STREET
CAMBRIDGE ST
NEW STREET
MARSHALL ST.
MARSHALL STREET
WEST STREET
CARNABY STREET
TYLER STREET
LITTLE MARLBOROUGH ST
KING STREET
CROSS STREET
SILVER STREET
HUSBAND STREET
PETER STREET
MEARDS STREET
OLD COMPTON STREET
CHURCH ST
KING ST
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
LITTLE PULTENEY STREET
BRIDLE STREET
GOLDEN SQUARE
UP JAMES ST
UP JOHN ST
WARWICK STREET
BEAK ST
NEW BURLINGTON STREET
OLD BURLINGTON MEWS
NEW BURLINGTON MEWS
BOYLE ST.
MILL ST
ARCHER STREET
RICHMOND ST
UP RUPERT ST
GERRARD ST
MACCLESFIELD ST.
PRINCE'S STREET

Real-world Data™ Saving Lives 1854

- Until the 19th century, it was not understood how disease transmission works (according to miasma theory, "bad air" was made responsible)
- During 1846-1860 cholera outbreak, physician John Snow hypothesized germ-contaminated water might be the actual cause of the disease
- Using a dot map of incidents, he found evidence that cholera cases are related to the public water pump on Broad Street
- 1st epidemiological study
- Germ theory was still rejected, until confirmed by more formal experiments (Louis Pasteur)





https://www.ph.ucla.edu/epi/snow/graphics/broadblock_original.jpg
https://en.wikipedia.org/wiki/John_Snow

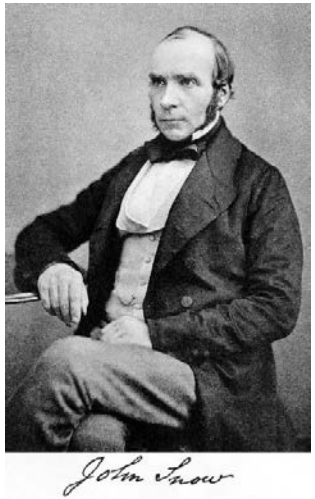
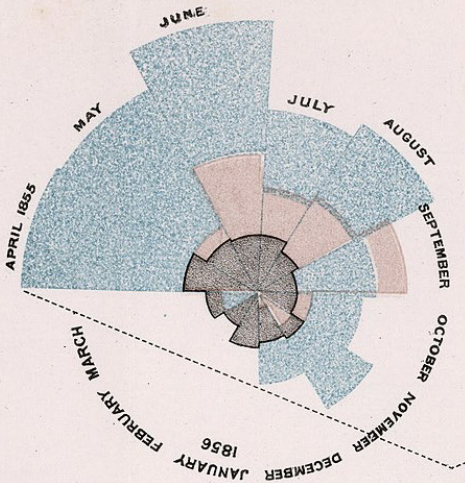
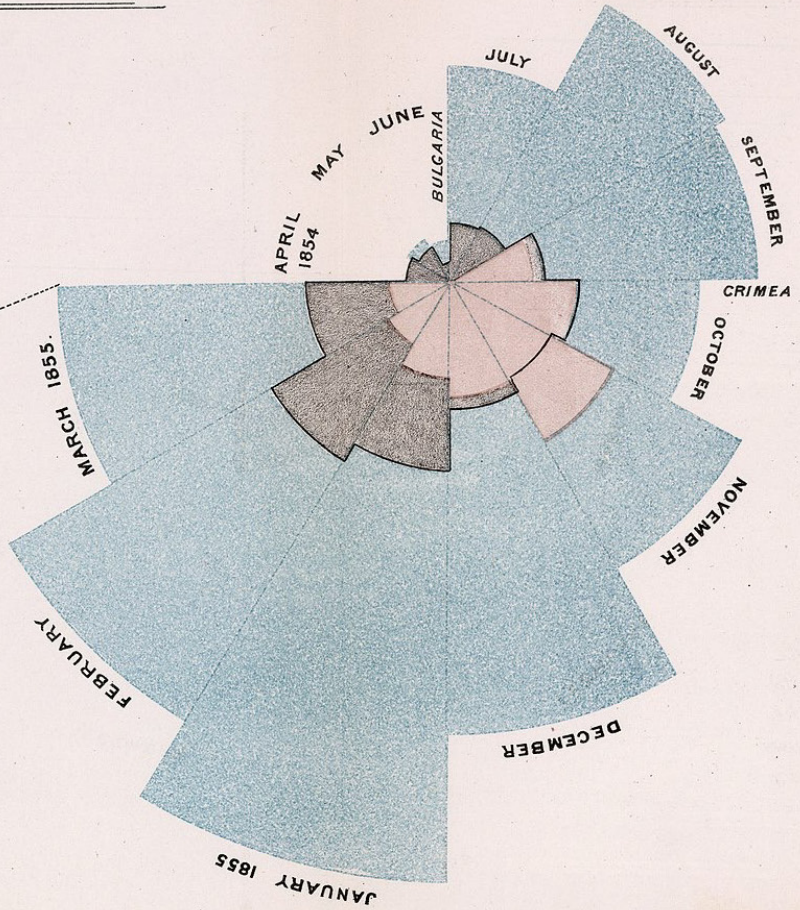


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Real-world Data™ Saving Lives 1858

- Florence Nightingale (1820-1910)
- Served as nursing administrator of British Army hospital in Scutari, Turkey during Crimean War (1853-1856)
- Rose / coxcomb / polar-area diagram designed by Florence Nightingale and William Farr, 1858
- Visualizes army deaths by reason
  - Blue: Preventable infectious diseases
  - Red: Wounds
  - Black: Other reasons

H. Lenthall, London

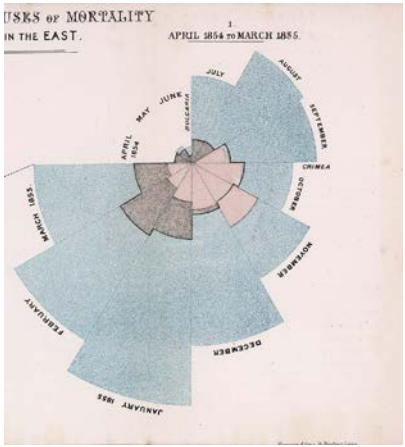
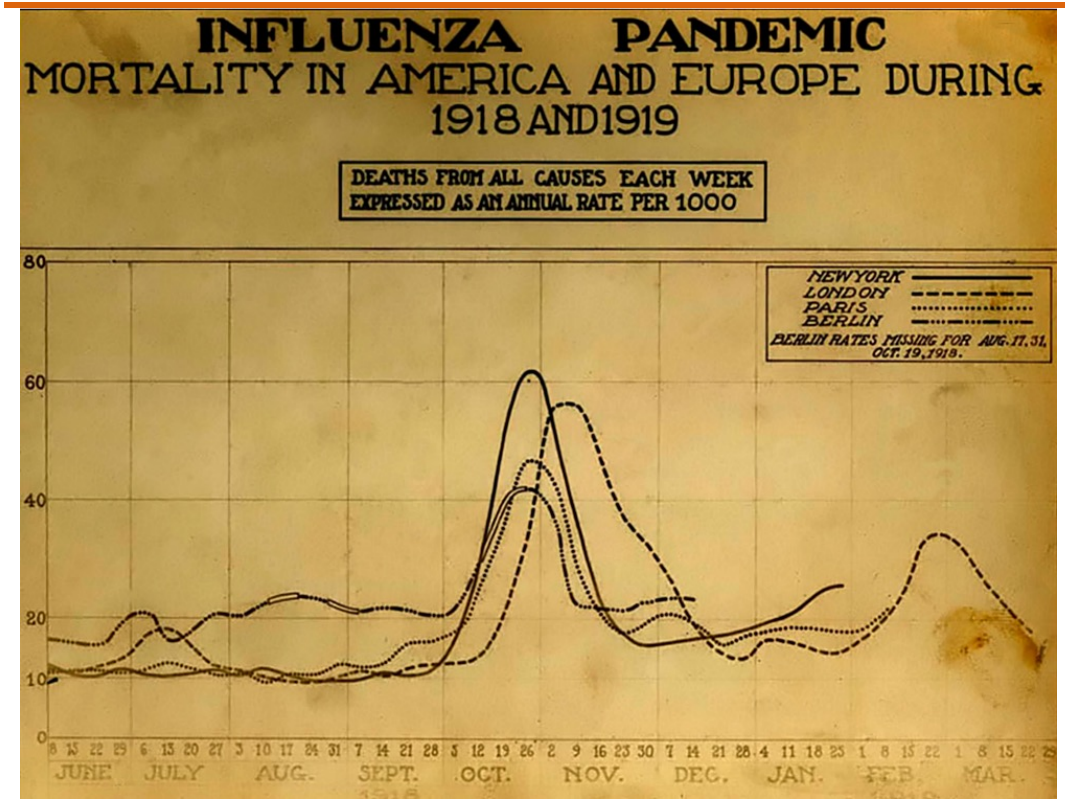


Diagram from 'Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army' by Florence Nightingale. Florence Nightingale Museum Collection

Influenza Pandemic (Spanish flu)



Pandemic Influenza: The Inside Story. Nicholls H, PLoS Biology Vol. 4/2/2006,

https://www.samerbergernachrichten.de/spanische-grippe-und-corona-werbeplakat/

2022 Situational Reports

- Official organizations create daily/weekly situational reports
- Most of the reports are static, e.g. HTML, Excel, PDF

COVID-19: Fallzahlen in Deutschland und weltweit

Fallzahlen in Deutschland

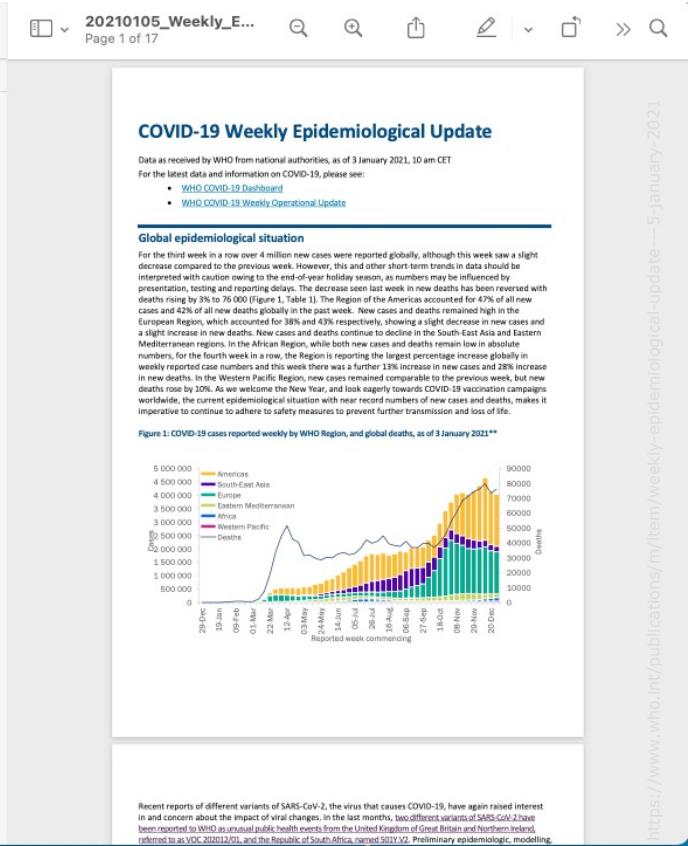
Stand: 6.1.2021, 00:00 Uhr (online aktualisiert um 08:25 Uhr)

Während der Weihnachtsfeiertage, zum Jahreswechsel und an den umgebenden Tagen ist bei der Interpretation der Fallzahlen zu beachten, dass zum einen meist weniger Personen einen Arzt aufsuchen, dadurch werden weniger Proben genommen und weniger Laboruntersuchungen durchgeführt. Dies führt dazu, dass weniger Erregernachweise an die zuständigen Gesundheitsämter gemeldet werden. Zum anderen kann es sein, dass nicht alle Gesundheitsämter und zuständigen Landesbehörden an allen Tagen an das RKI übermitteln.

Die Anzahl der Fälle - und deren Differenz zum Vortag - und die Anzahl der Todesfälle beziehen sich auf Fälle, die dem RKI täglich übermittelt werden. Dies beinhaltet Fälle, die am gleichen Tag oder bereits an früheren Tagen an das Gesundheitsamt gemeldet worden sind. Bei den Fällen in den letzten 7 Tagen und der 7-Tage-Inzidenz liegt das Meldedatum beim Gesundheitsamt zugrunde, also das Datum, an dem das lokale Gesundheitsamt Kenntnis über den Fall erlangt und ihn elektronisch erfasst hat.

Bundesland	Elektronisch übermittelte Fälle				
	Anzahl	Diffe-renz zum Vortag	Fälle in den letzten 7 Tagen	7-Tage-Inzidenz	Todesfälle
Baden-Württemberg	249.397	2.957	13.769	124	5.180
Bayern	338.786	1.691	18.224	139	7.260
Berlin	100.760	1.235	4.287	117	1.390
Brandenburg	44.556	1.393	3.626	144	1.083
Bremen	13.915	112	502	74	209
Hamburg	38.575	696	2.085	113	695
Hessen	142.925	1.913	8.161	130	3.211
Meck-	13.225	487	1.551	96	200

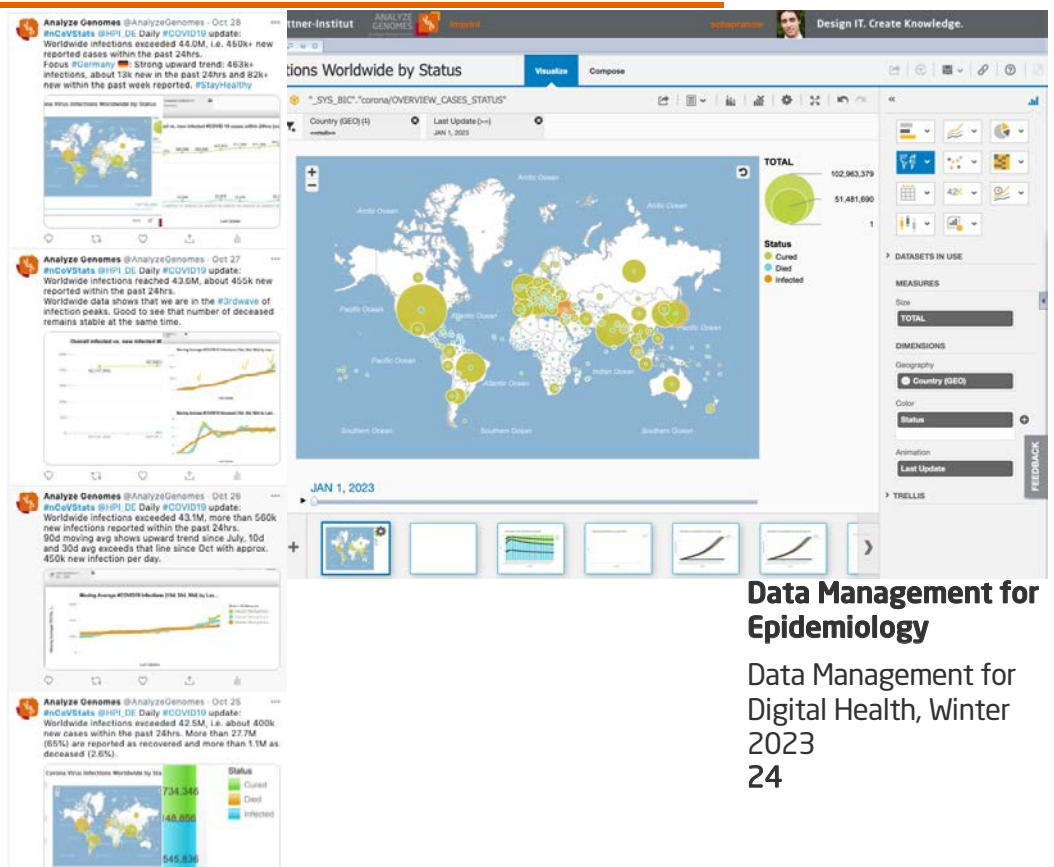
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html



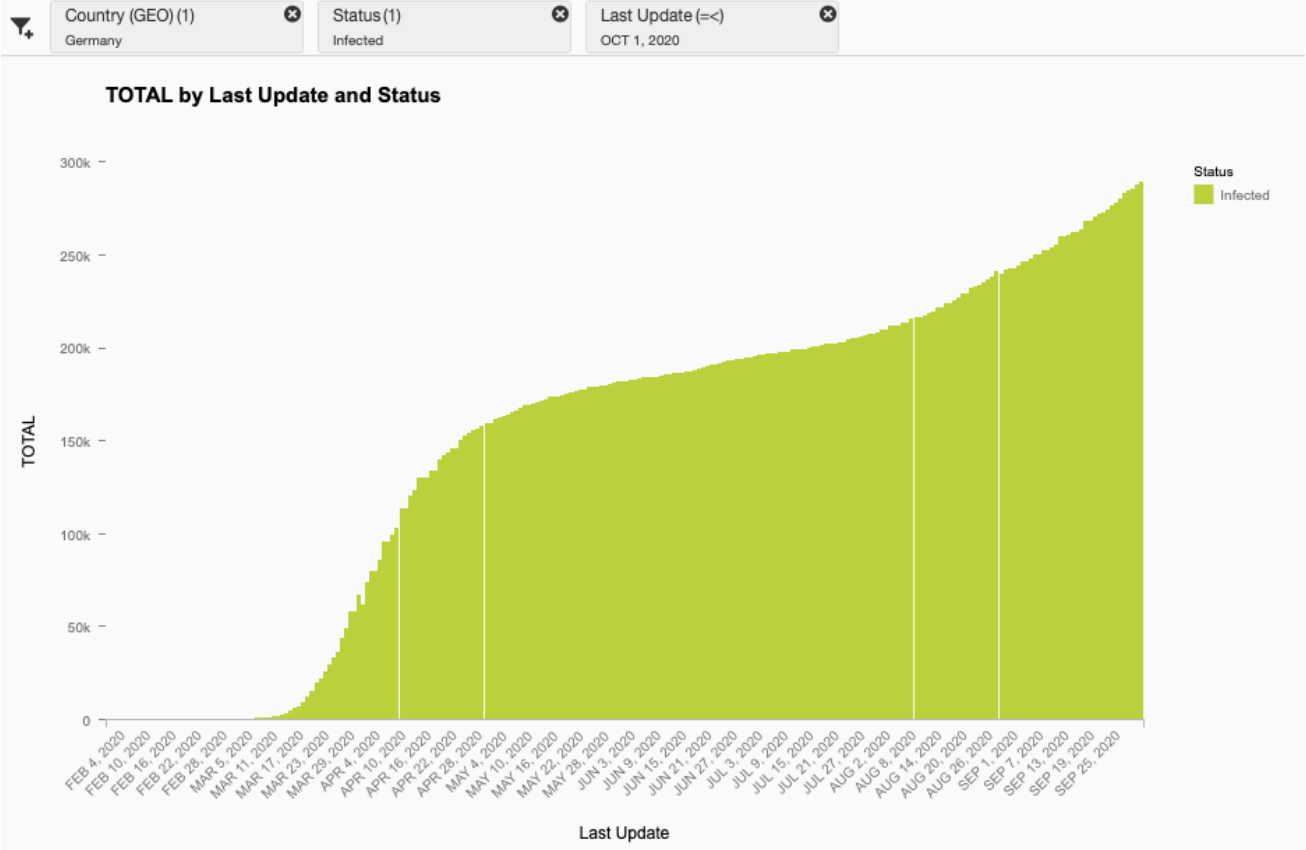
https://www.who.int/publications/m/item/weekly-epidemiological-update---5-january-2021

How we have started #nCoVStats

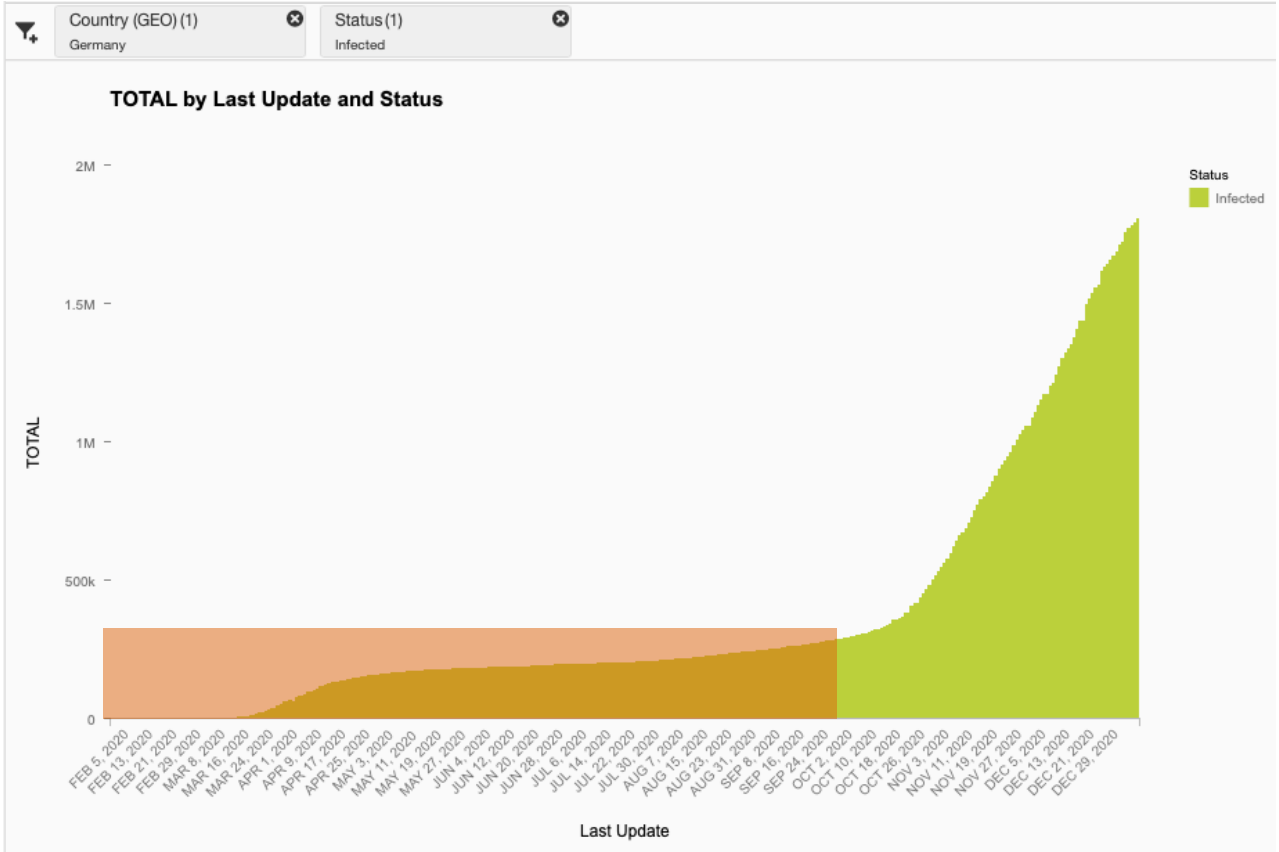
- Kickstarted in Jan, launched Feb 1, 2020
- Incorporates in-memory database to enable real-time analytics
- Longitudinal pandemic database with 220k+ entries for 680+ regions
- Worldwide open access to latest data
- Interactive graphical data analysis tool
- Daily #nCoVStats update on Twitter
- Link: https://we.analyzegenomes.com/apps/nCoVStats/



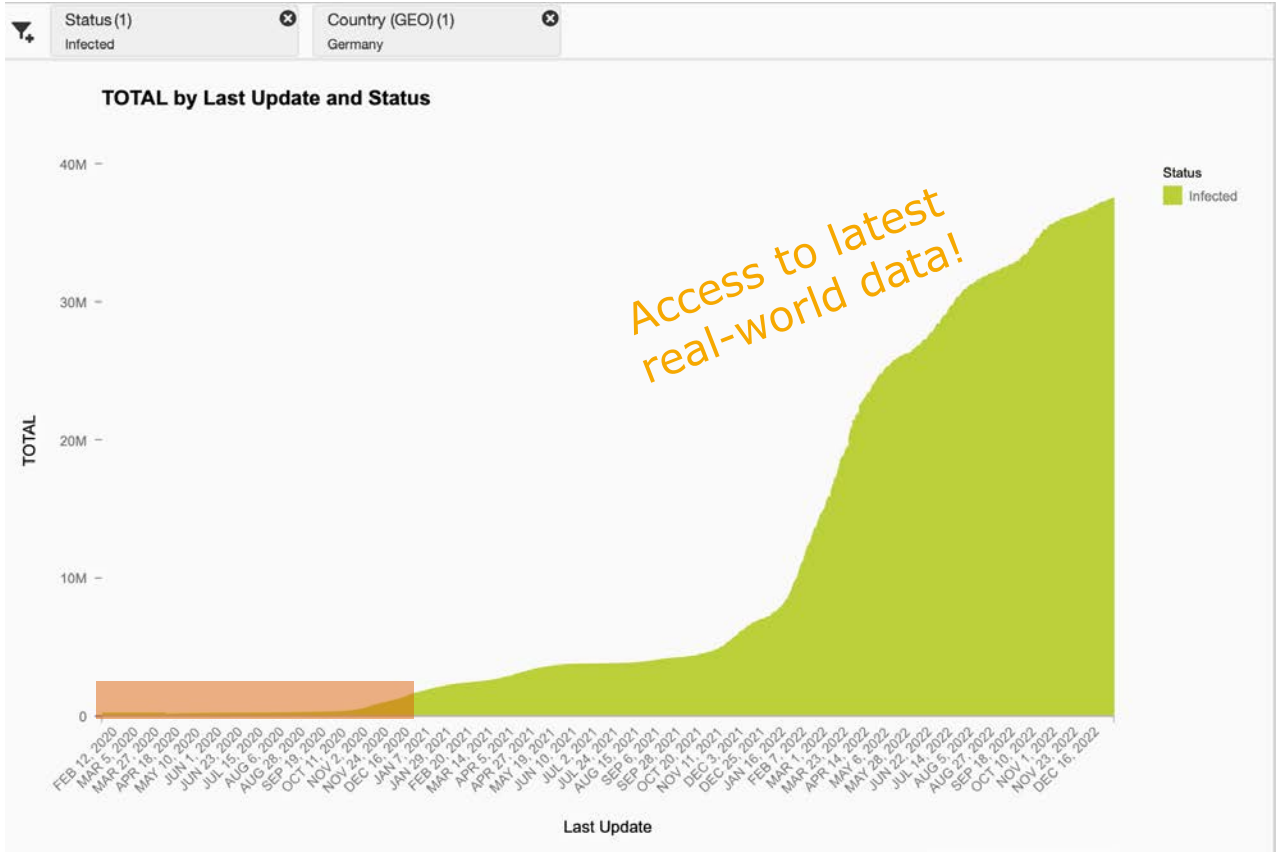
What is Crucial for Informed Decision Making?



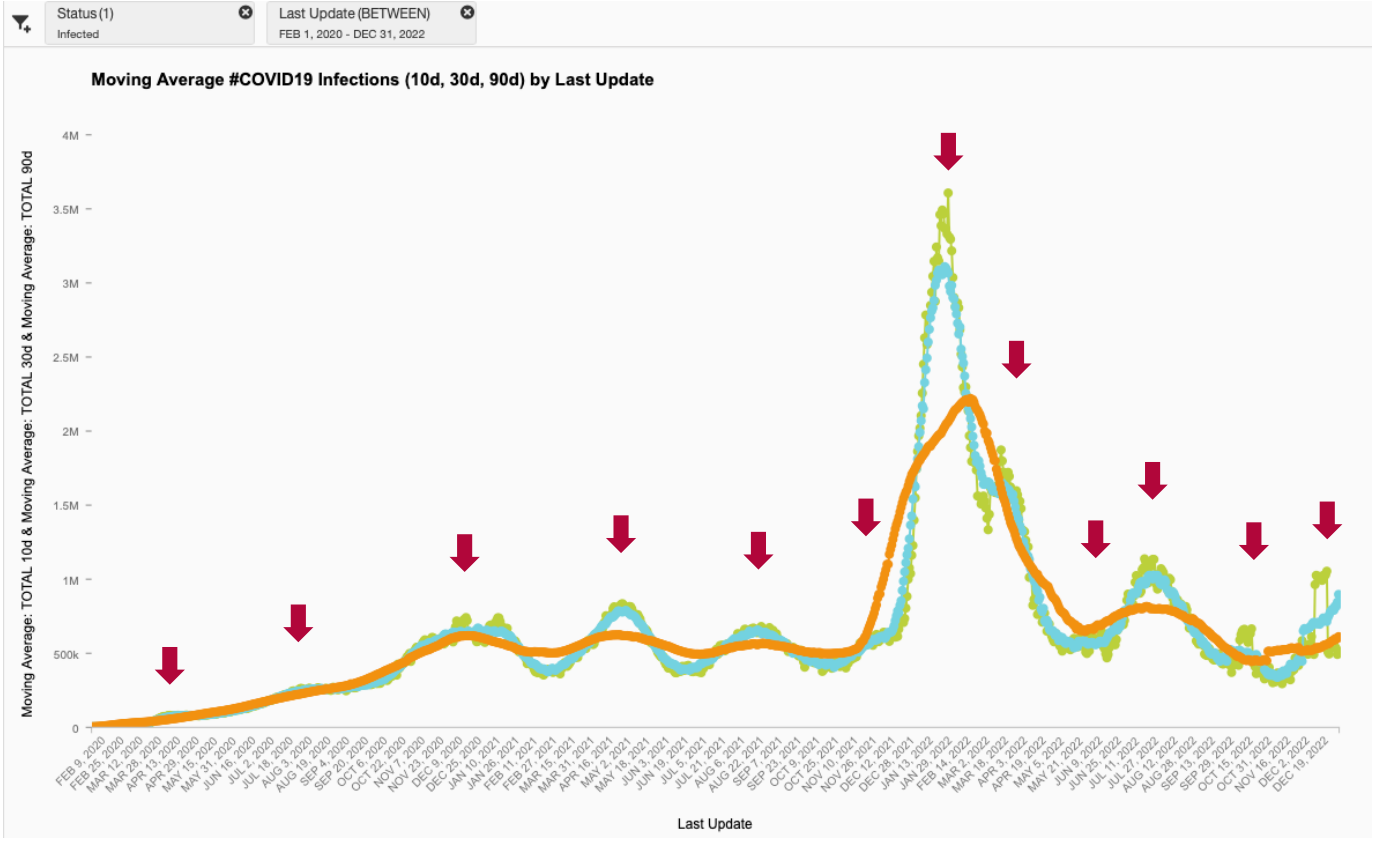
What is Crucial for Informed Decision Making? (cont'd)



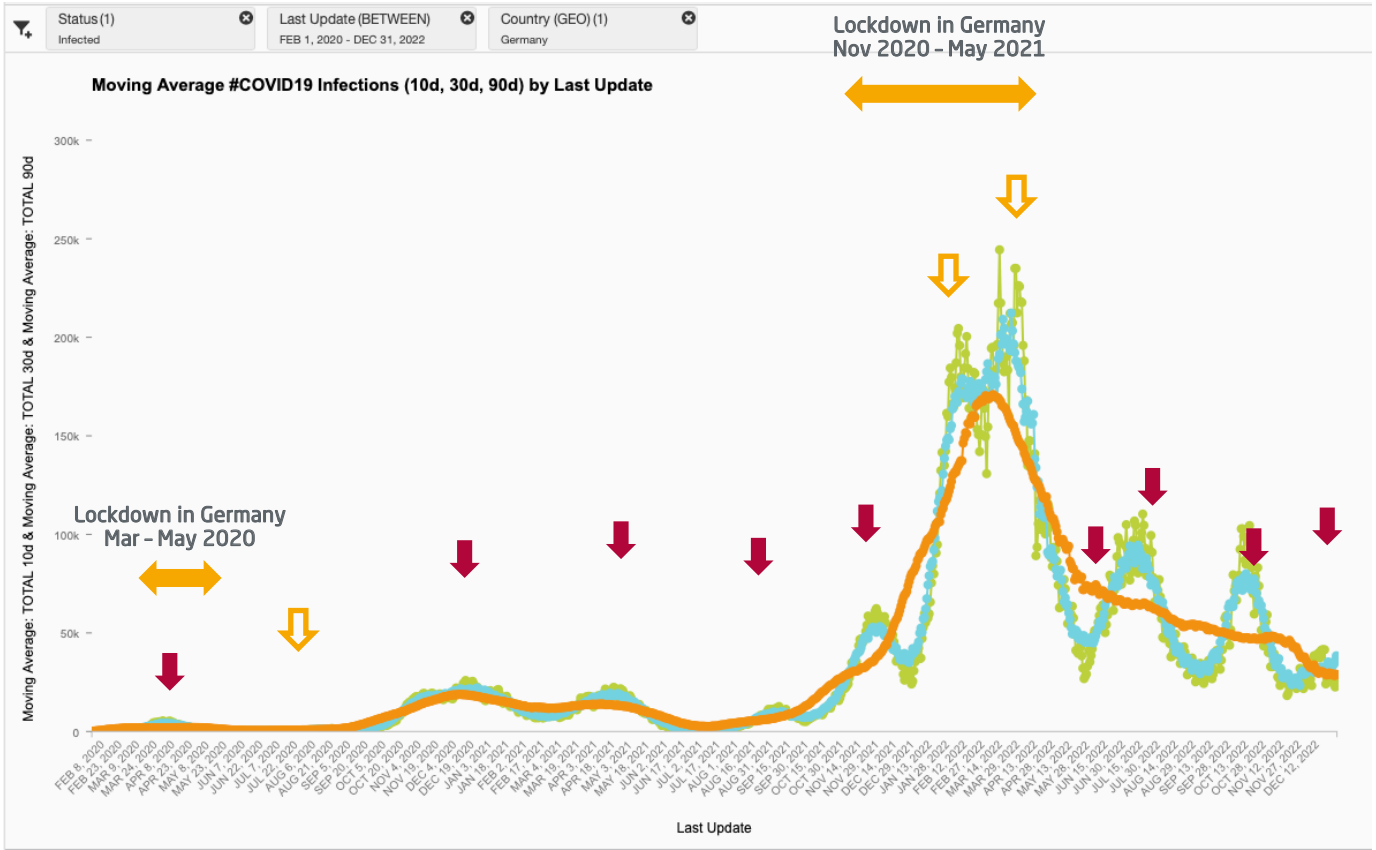
What is Crucial for Informed Decision Making? (cont'd)



New Infections (Worldwide) over Time (Feb 2020-Dec 2022)



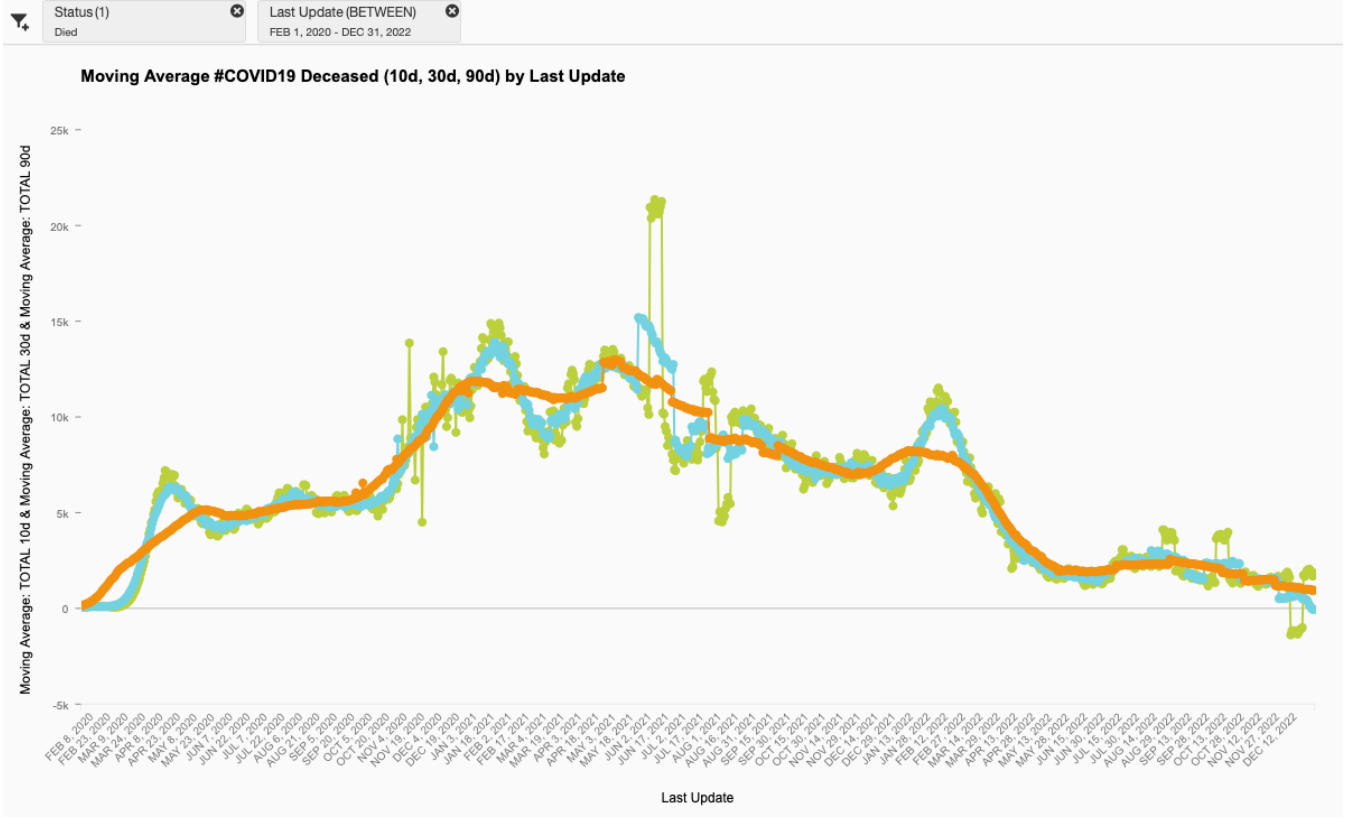
New Infections (Germany) over Time (Feb 2020-Dec 2022)



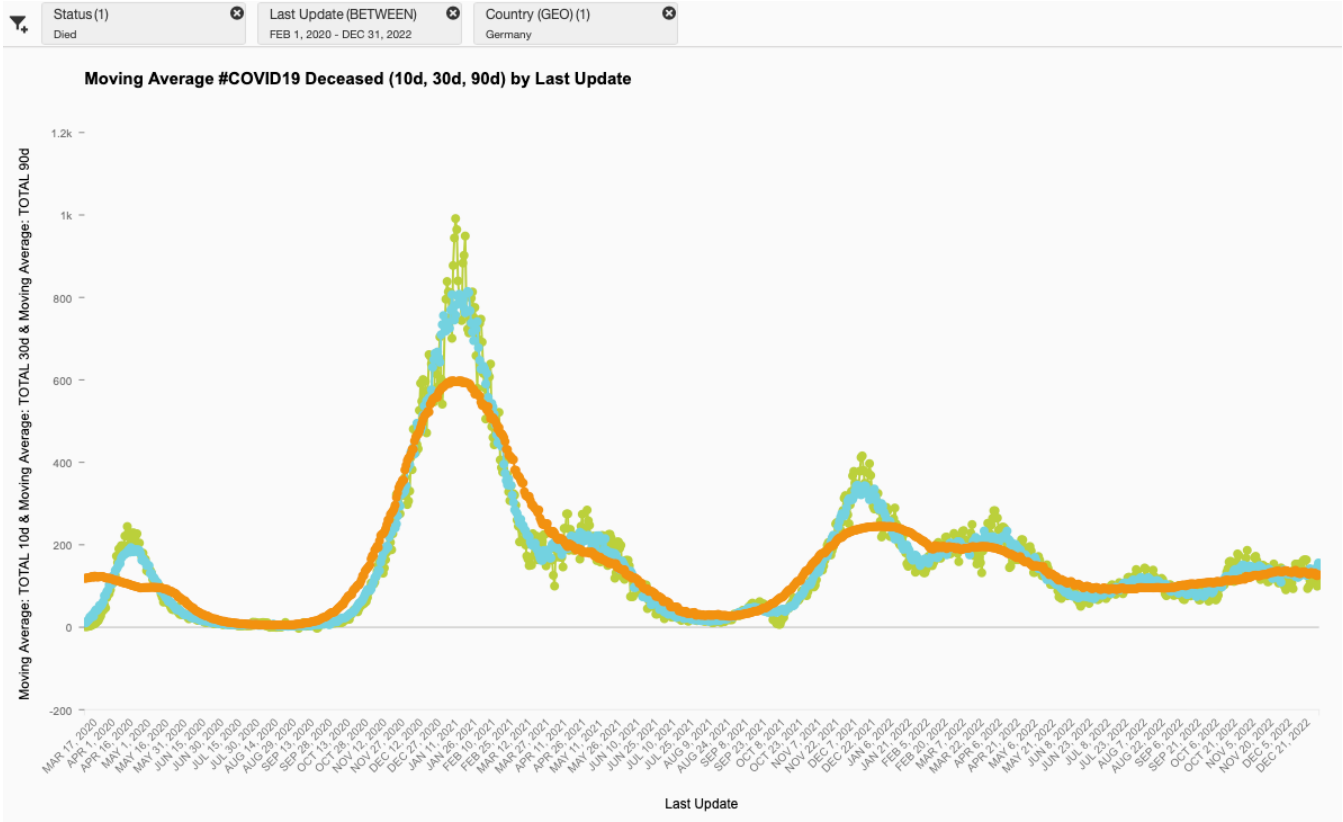
Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

New Deceased (Worldwide) over Time (Feb 2020-Dec 2022)



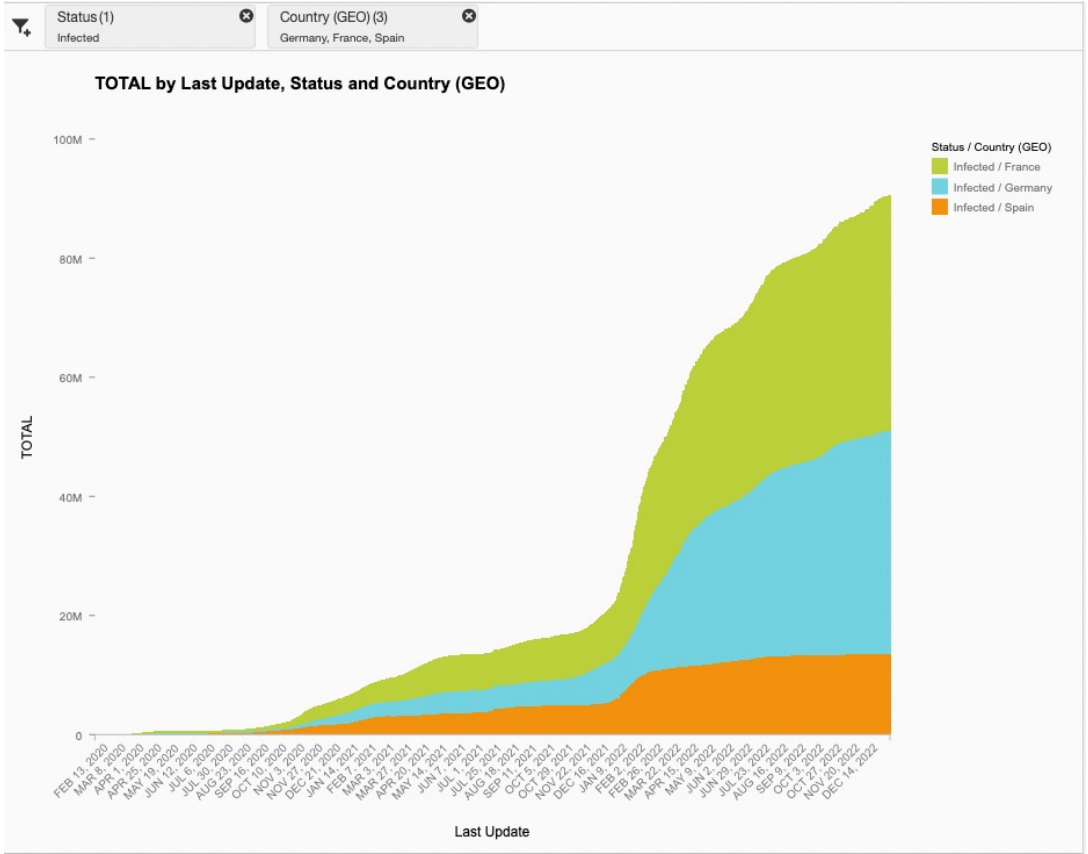
New Deceased (Germany) over Time (Feb 2020-Dec 2022)



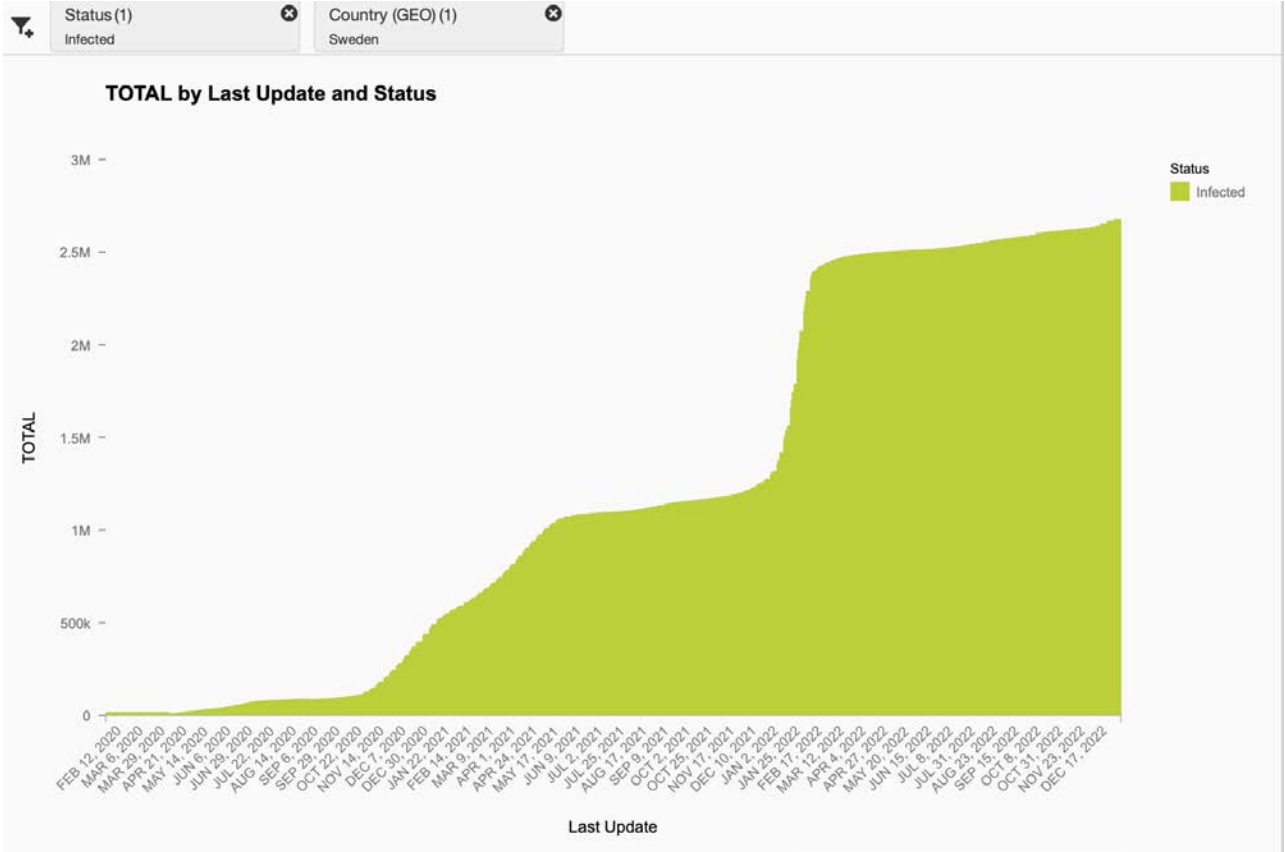
Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

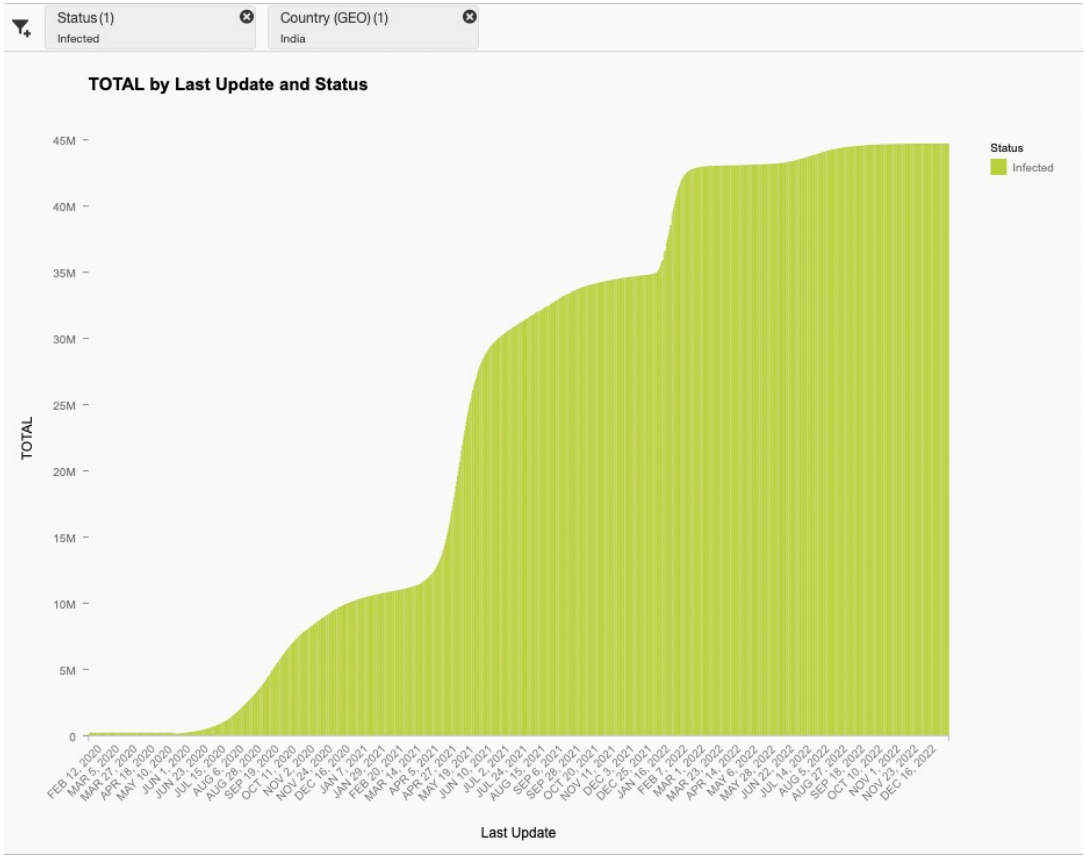
Aggregated Infections (France, Germany, Spain) over Time (Feb 2020-Dec 2022)

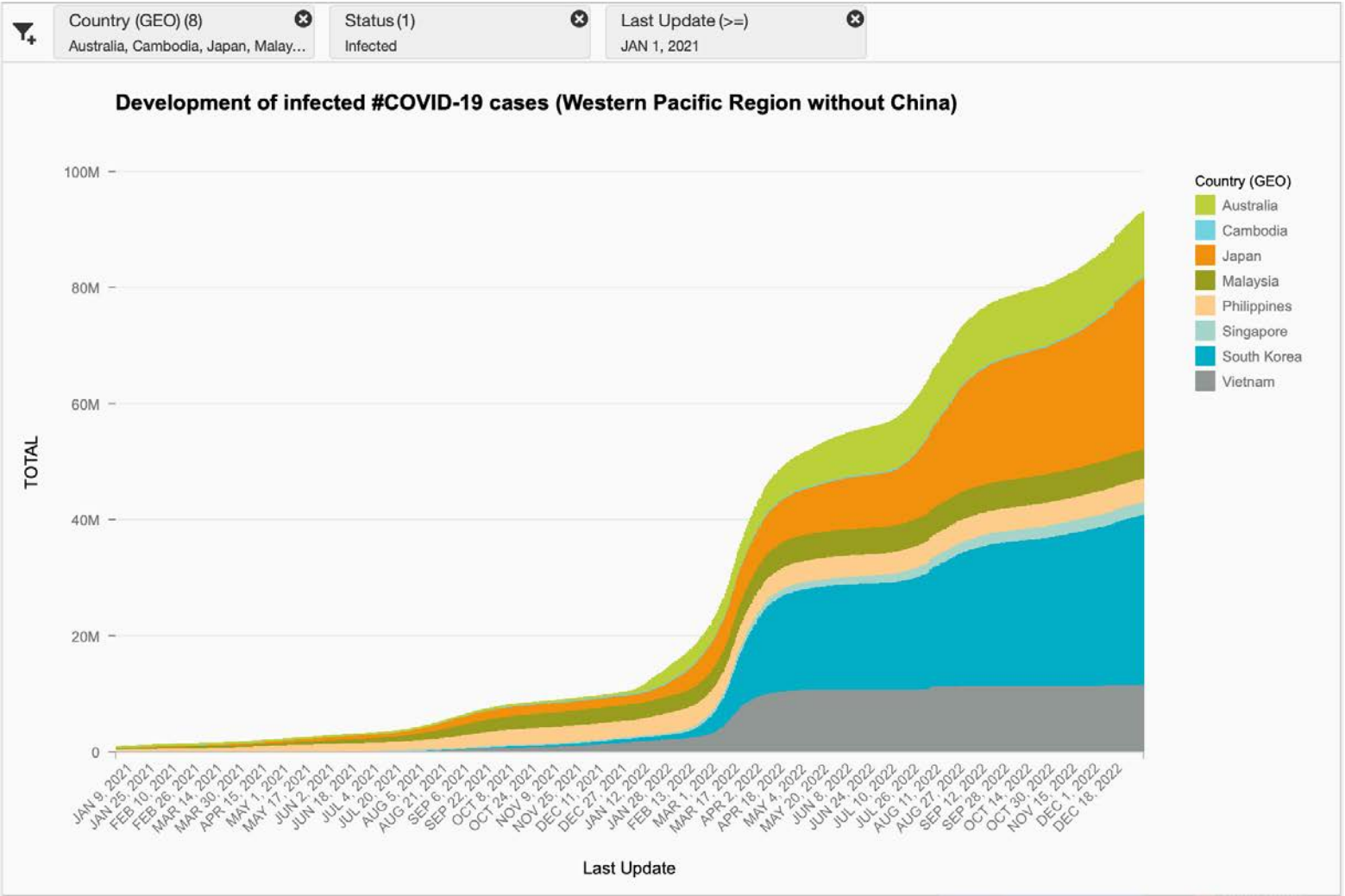


Aggregated Infections (Sweden) over Time (Feb 2020-Dec 2022)



Aggregated Infections (India) over Time (Feb 2020-Dec 2022)

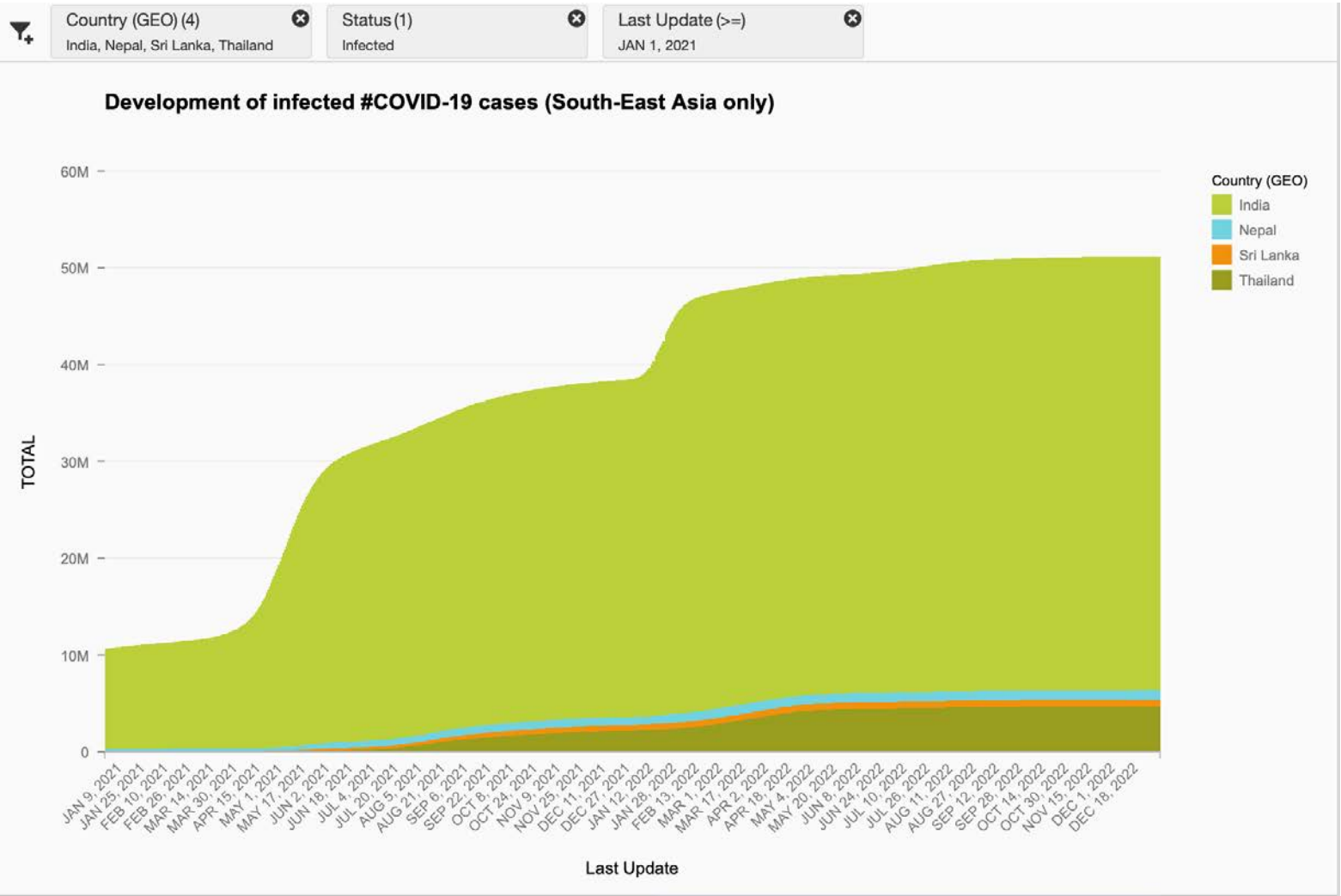




Hasso Plattner Institut

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023



Hasso Plattner Institut

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

#nCoVStats: Data Sources

- Primary data source for international reports
- Real-time updated case numbers
- Human-readable, but also crawlable

地区	确诊	死亡	治愈
湖北	31728	974	2277
广东	1177	1	211
浙江	1117		267
河南	1105	7	215
湖南	912	1	244
安徽	860	4	105
江西	804	1	127
江苏	515		93
重庆	489	2	72
山东	487	1	80
四川	417	1	85
黑龙江	360	8	28
北京	342	3	48
上海	303	1	52
福建	267		43

https://ncov.dxy.cn/ncovh5/view/pneumonia

#nCoVStats: Data Ingestion

```
for i in 1 2
do
        ID="getListByCountryTypeService$i"
        echo $CONTENT | grep 'script id=\"'$ID'[a-z]*\"' | sed -E 's/^.*script id=\"'$ID'[a-z]*\">([^>]+)<\/script>.*$/\1/g' | sed -E 's/try \{(.*)\}catch\(e\)\{\}/\1/g' | sed -E 's/^[^\[]+(.*)$/\1/g' | jq -r '.[] | .modifyTime |= ((. // now*1000) / 1000 | strftime("%Y-%m-%d")) | [.provinceName, .confirmedCount, .deadCount, .modifyTime, .curedCount] | @csv' >> $TMP_FILE
done
```

- Crawler checks regularly data sources for updated numbers
- Downloads web page and filters numbers
- Creates CSV for database import
- Triggers IMDB import of updated case numbers
- Data quality?

```
"美国",2510323,125539,"2020-06-28",679308
"巴西",1315941,57103,"2020-06-28",715905
"英国",310250,43514,"2020-06-28",539
"俄罗斯",627646,8969,"2020-06-28",393352
"智利",267766,5347,"2020-06-28",44946
"印度",528859,16095,"2020-06-28",309712
"巴基斯坦",200832,4073,"2020-06-28",86906
"秘鲁",275989,9135,"2020-06-28",164024
"孟加拉国",133978,1695,"2020-06-28",54318
"西班牙",248469,28341,"2020-06-28",150376
"南非",131800,2413,"2020-06-28",67094
"法国",162936,29778,"2020-06-28",75649
"瑞典",65484,5304,"2020-06-28",4971
"沙特阿拉伯",178504,1511,"2020-06-28",122127
"哥伦比亚",88591,2939,"2020-06-28",36273
"厄瓜多尔",53856,4406,"2020-06-28",5040
"埃及",63923,2708,"2020-06-28",17140
"阿根廷",57744,1207,"2020-06-28",13576
"墨西哥",212802,26381,"2020-06-28",149318
"比利时",61209,9732,"2020-06-28",16858
"荷兰",50074,6105,"2020-06-28",11868
"伊朗",220180,10364,"2020-06-28",180661
"加拿大",103032,8516,"2020-06-28",65973
"/tmp/cases.csv" 215L, 8435C
```

```
echo "IMPORT FROM CSV FILE '"$TMP_FILE"' INTO \"CORONA2020\".\"CASES\" WITH RECORD DELIMITED BY '\n' FIELD DELIMITED BY ',' OPTIONALLY ENCLOSED BY '\"' ERROR LOG '"$TMP_FILE.err"';" | hdbsql
```

#nCoVStats: Defining Analysis Reports

- Jan 6, 2021

184 rows retrieved - 116 ms

Chart | Table | Grid | HTML

NAME_EN	Population_2020	INFECT	INFECTED_BY_POP	CURED	CURED_BY_POP
United States	331.002.651	21.044.020	0,064	9.550.165	0,029
India	1.380.004.385	10.356.844	0,008	9.975.958	0,007
Brazil	212.559.417	7.812.007	0,037	6.963.407	0,033
Russia	145.934.462	3.284.384	0,023	2.662.668	0,018
United Kingdom	67.886.011	2.774.479	0,041	2.473	0
France	65.273.511	2.680.239	0,041	197.503	0,003
Turkey	84.339.067	2.270.101	0,027	2.155.338	0,026
Italy	60.461.826	2.181.619	0,036	1.536.129	0,025
Spain	46.754.778	1.982.544	0,042	151.294	0,003
Germany	83.783.942	1.804.286	0,022	1.424.259	0,017
Colombia	50.882.891	1.702.966	0,033	1.559.010	0,031
Argentina	45.195.774	1.662.713	0,037	1.466.601	0,032
Mexico	128.932.753	1.455.219	0,011	1.105.834	0,009
Poland	37.846.611	1.330.543	0,035	1.078.892	0,029
Iran	83.992.949	1.255.620	0,015	8.123.231	0,097
South Africa	59.308.690	1.127.759	0,019	980.879	0,017

https://we.analyzegenomes.com/apps/ncovstats/

#nCoVStats: Defining Analysis Reports

- Jan 1, 2023

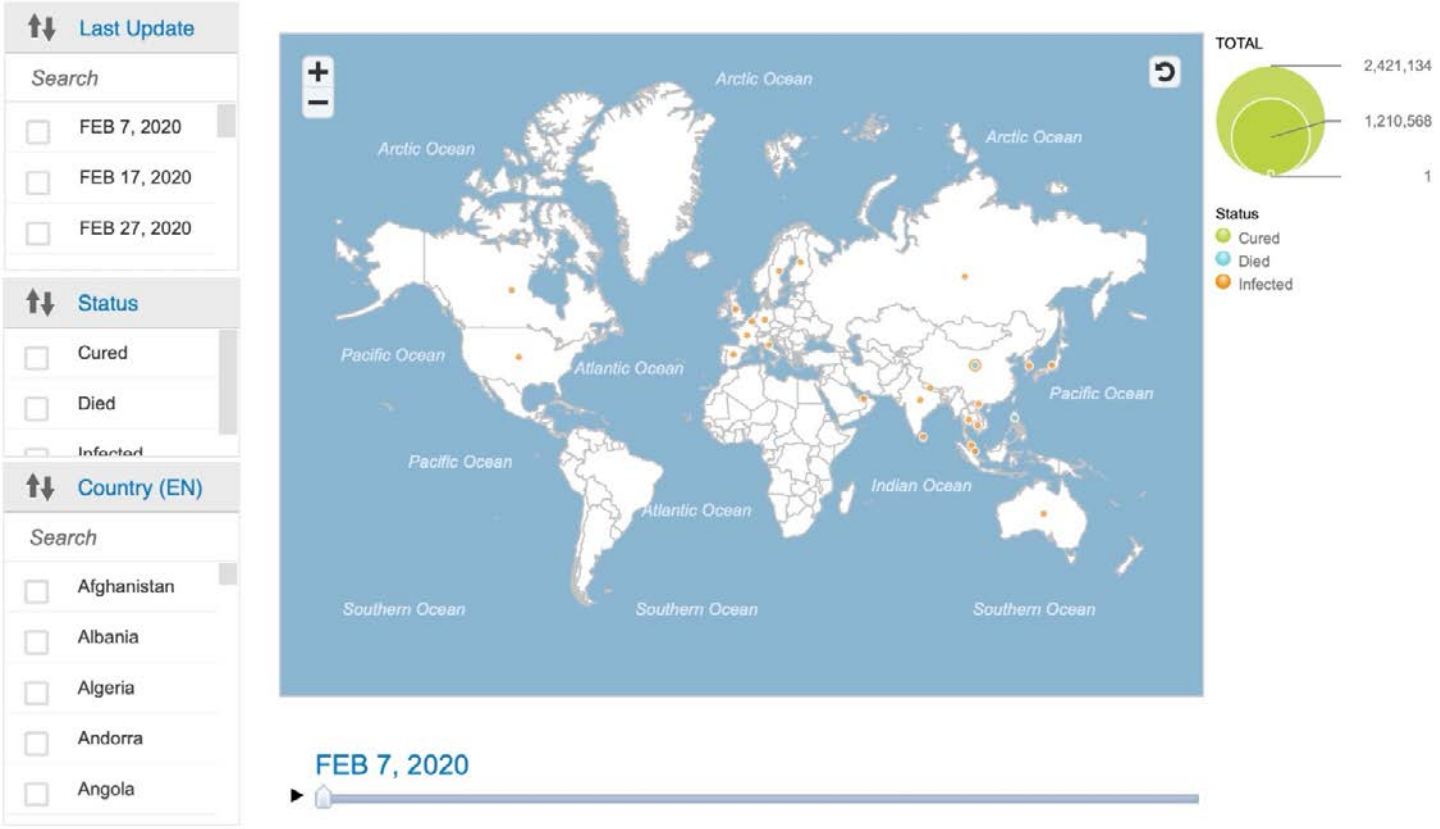
200 rows retrieved - 127 ms

Chart | Table | Grid | HTML

NAME_EN	Population_2	INFECTED	INFECTED_BY_POP	CURED_	CURED_BY_POP_	DIED_	DIED_BY_POP_
United States	331.002.651	102.513.690	0,31	99.513.507	0,301	1.117.983	0,003
India	1.380.004.385	44.679.564	0,032	44.144.029	0,032	530.702	0
France	65.273.511	39.316.017	0,602	38.342.881	0,587	161.962	0,002
Germany	83.783.942	37.369.865	0,446	36.615.400	0,437	161.465	0,002
Brazil	212.559.417	36.354.255	0,171	34.938.186	0,164	693.941	0,003
Japan	126.476.461	29.212.535	0,231	21.105.754	0,167	57.266	0
South Korea	51.269.185	29.116.800	0,568	27.893.416	0,544	32.219	0,001
Italy	60.461.826	25.143.705	0,416	24.541.402	0,406	184.642	0,003
United Kingdom	67.886.011	24.135.084	0,356	23.844.243	0,351	198.937	0,003
Russia	145.934.462	21.798.509	0,149	21.207.802	0,145	393.712	0,003
Turkey	84.339.067	17.042.722	0,202	<Null>	<Null>	101.492	0,001
Spain	46.754.778	13.684.258	0,293	13.486.683	0,288	117.095	0,003
Australia	25.499.884	11.131.707	0,437	10.979.282	0,431	17.052	0,001
Argentina	45.195.774	9.891.139	0,219	9.609.732	0,213	130.124	0,003
China	1.439.323.776	8.847.360	0,006	8.454.934	0,006	15.253	0
Netherlands	17.134.872	8.569.228	0,5	8.520.849	0,497	22.989	0,001
Iran	83.992.949	7.561.140	0,09	7.336.146	0,087	144.685	0,002
Mexico	128.932.753	7.234.467	0,056	6.468.257	0,05	331.099	0,003
Indonesia	273.523.615	6.719.815	0,025	6.549.332	0,024	160.612	0,001
Poland	37.846.611	6.368.479	0,168	5.335.940	0,141	118.533	0,003

https://weanalyzegenomes.com/apps/ncovstats/

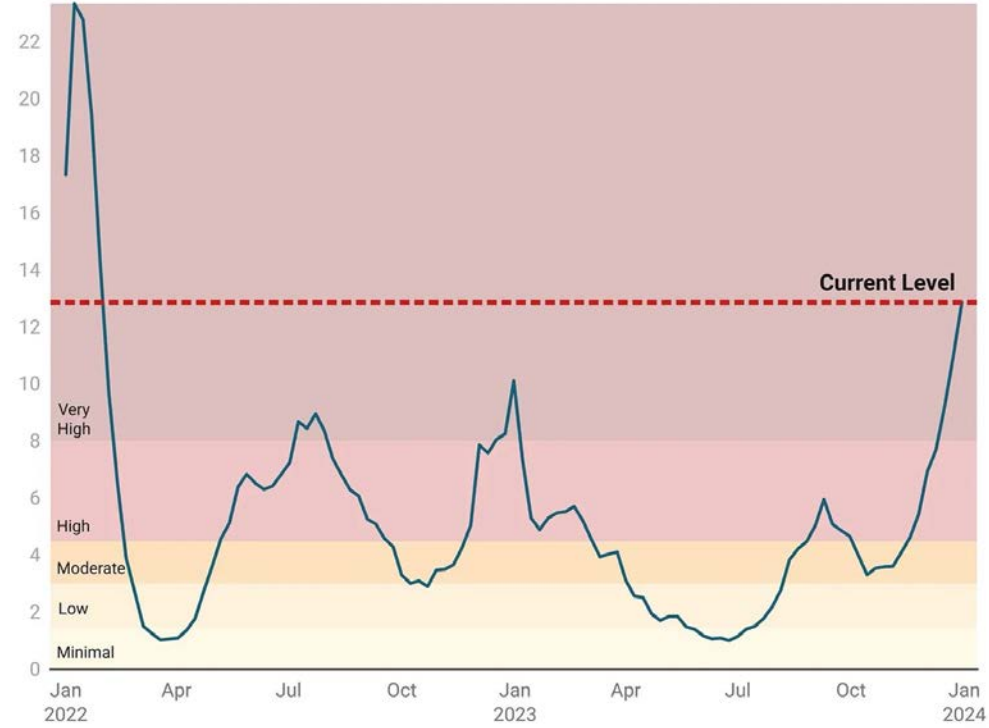
#nCoVStats: Defining Analysis Reports



Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

Excurse: COVID-19 Wastewater Levels in the U.S. (2024)



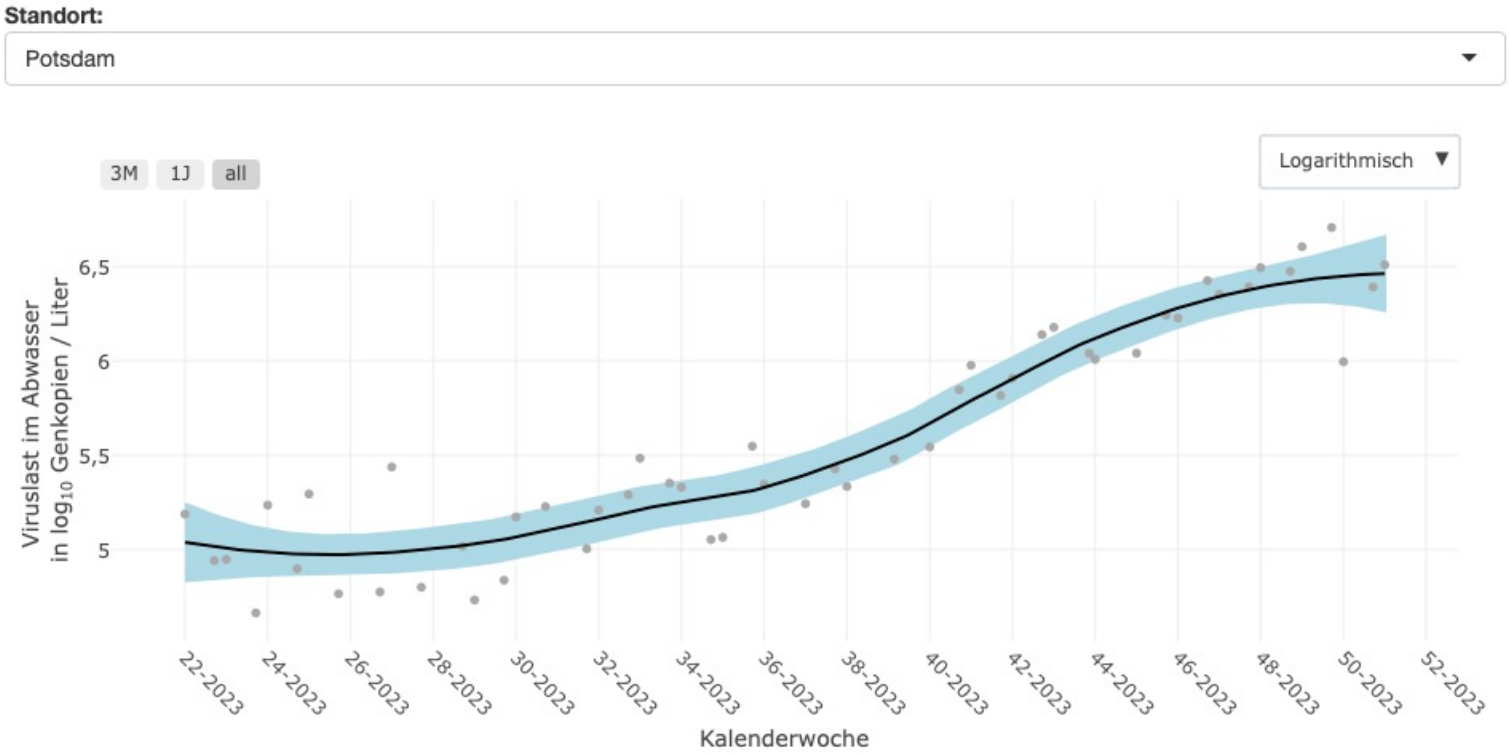
Data last updated 2024-01-04

Chart: @luckytran • Source: CDC • Created with Datawrapper

COVID-19 Wastewater Levels in the U.S. (Jan 4, 2024)

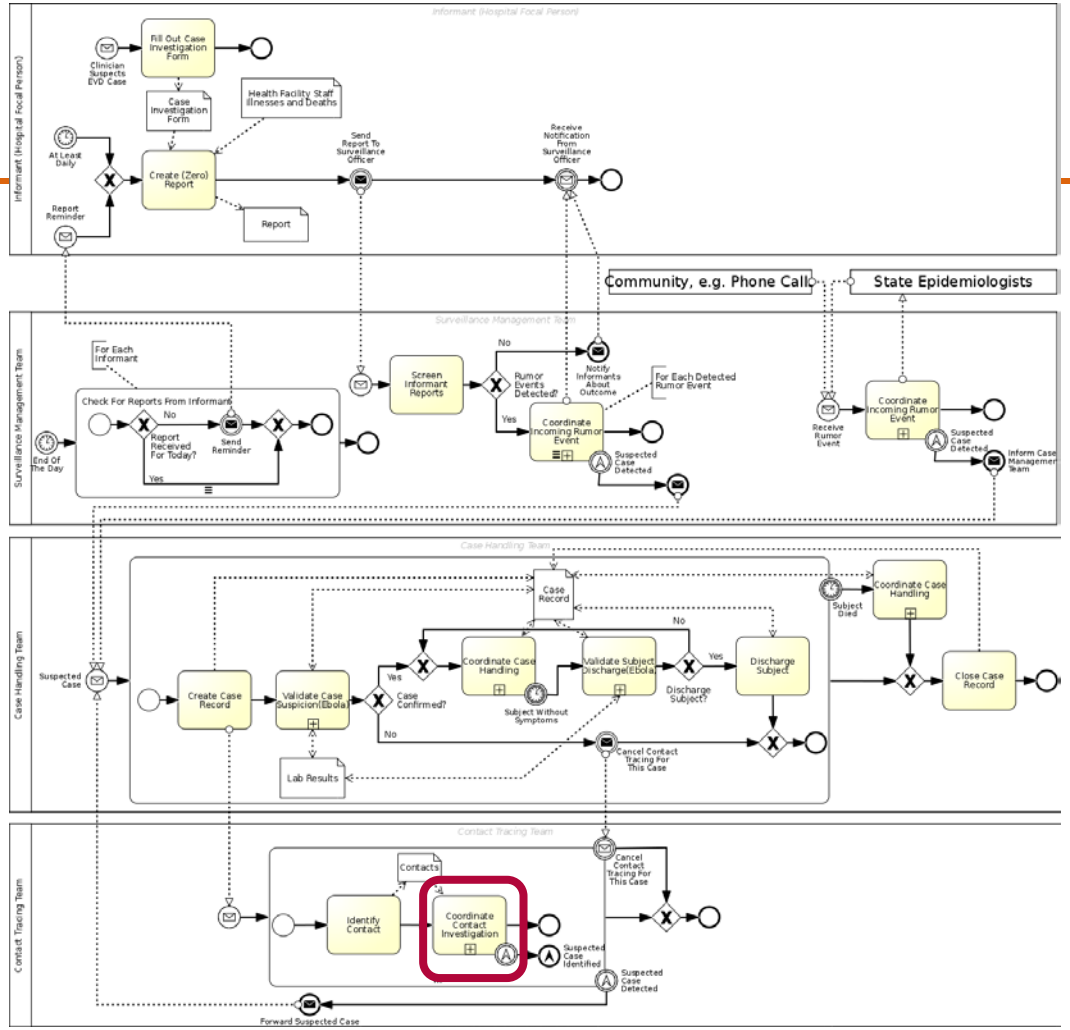
Chart shows national trends of SARS-COV-2 viral activity levels in U.S. wastewater. The vertical axis shows the Wastewater Viral Activity Level, which indicates changes in SARS-CoV-2 virus levels in wastewater compared to the baseline level (in standard deviations).

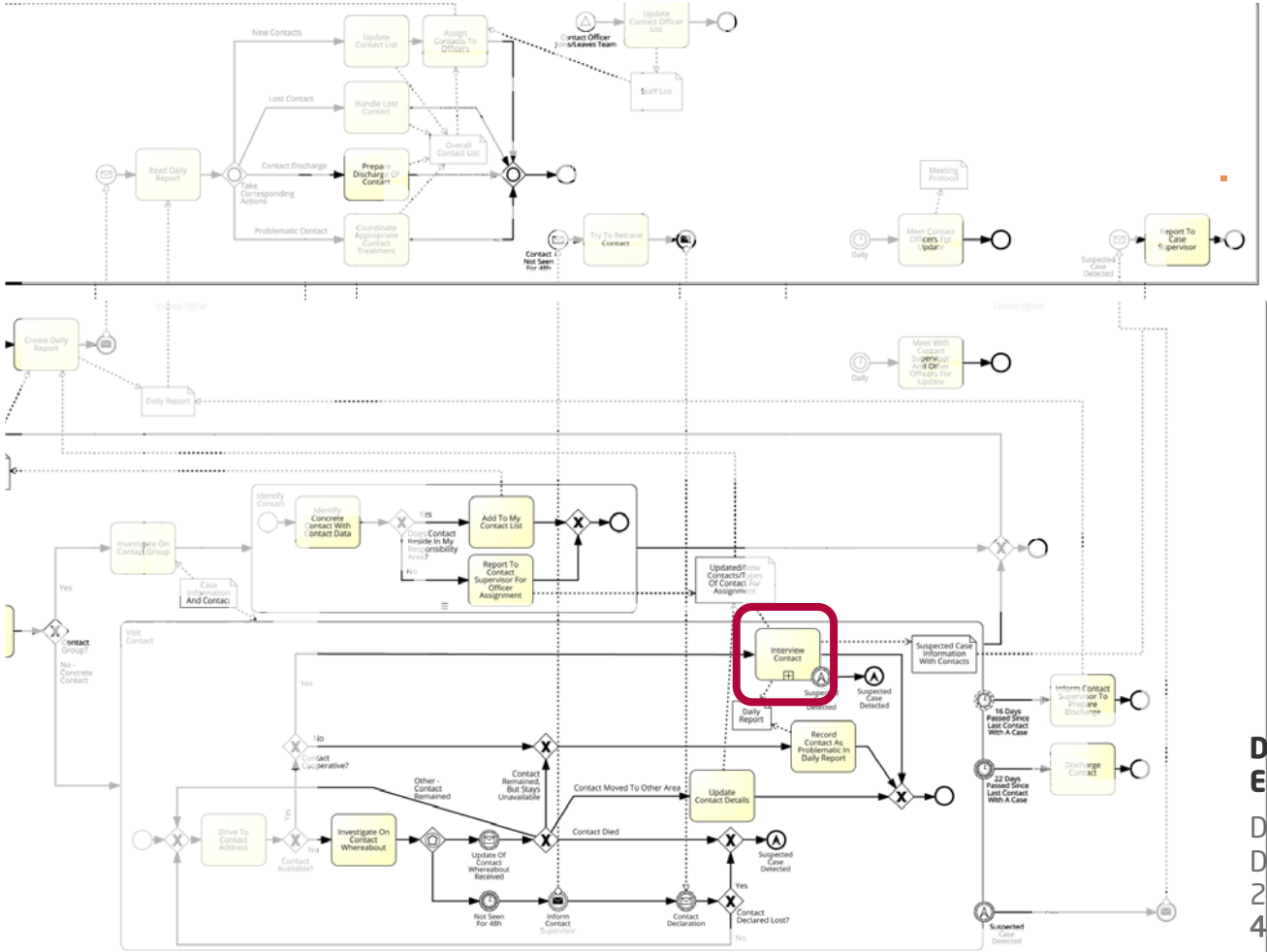
Excurse: COVID-19 Wastewater Levels in Potsdam (2024)



SORMAS App for Contact Tracing by HPI: Field Test in Nigeria during Ebola Epidemic 2014/15

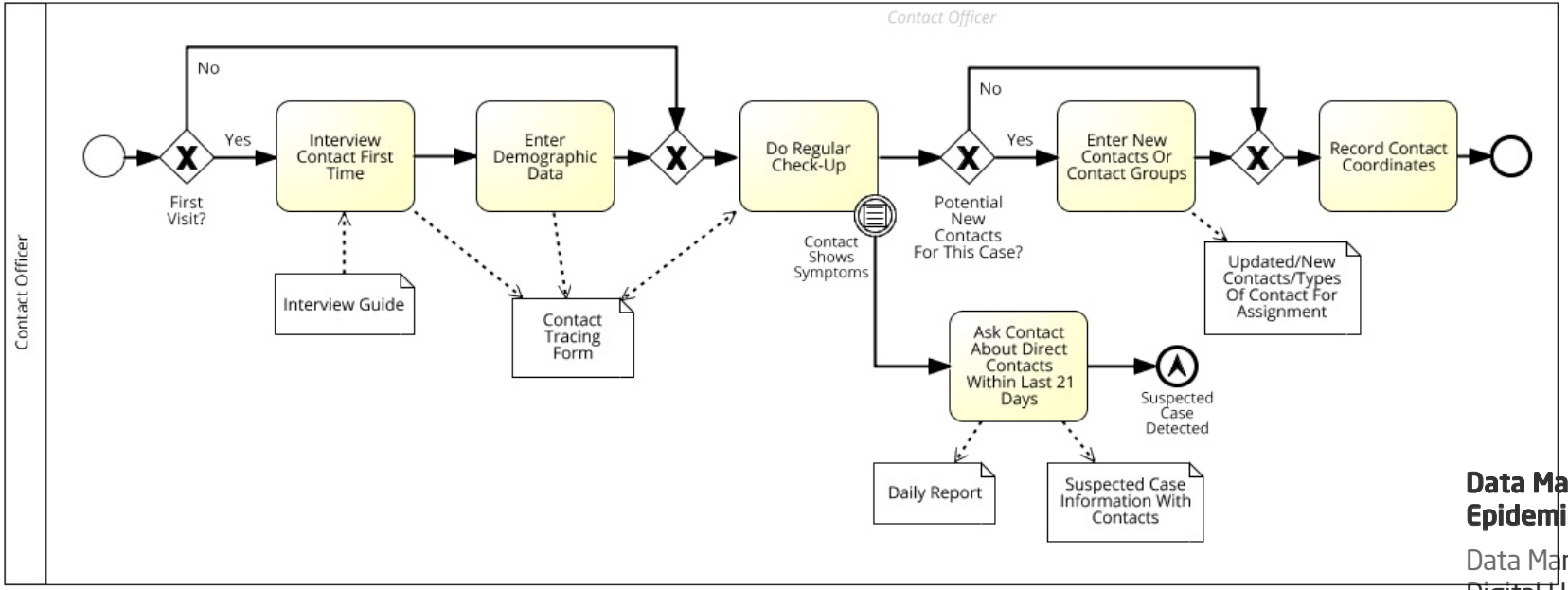
- Contact tracing app developed by HPI students together with leading experts.
- Requirements:
  - Fast training of contact tracers, ease of use
  - Work with commodity Android smartphones
  - Support real-time response in case of reported infection


Informant (Hospital Focal Person)
Fill Out Case Investigation Form
Clinician Suspects EVD Case
Case Investigation Form
Health Facility Staff Illnesses and Deaths
At Least Daily
Create (Zero) Report
Report
Report Reminder
Send Report To Surveillance Officer
Receive Notification From Surveillance Officer
Community, e.g. Phone Call
State Epidemiologists
Surveillance Management Team
For Each Informant
Check For Reports From Informant
No
Report Received For Today?
Send Reminder
Yes
End Of The Day
Screen Informant Reports
No
Rumor Events Detected?
Yes
Notify Informants About Outcome
For Each Detected Rumor Event
Coordinate Incoming Rumor Event
Suspected Case Detected
Receive Rumor Event
Coordinate Incoming Rumor Event
Suspected Case Detected
Inform Case Management Team
Case Handling Team
Suspected Case
Create Case Record
Validate Case Suspicion(Ebola)
Case Confirmed?
Yes
No
Lab Results
Case Record
Coordinate Case Handling
Subject Without Symptoms
Validate Subject Discharge(Ebola)
Discharge Subject?
Yes
No
Discharge Subject
Cancel Contact Tracing For This Case
Subject Died
Coordinate Case Handling
Close Case Record
Contact Tracing Team
Contacts
Identify Contact
Coordinate Contact Investigation
Suspected Case Identified
Suspected Case Detected
Cancel Contact Tracing For This Case
Forward Suspected Case



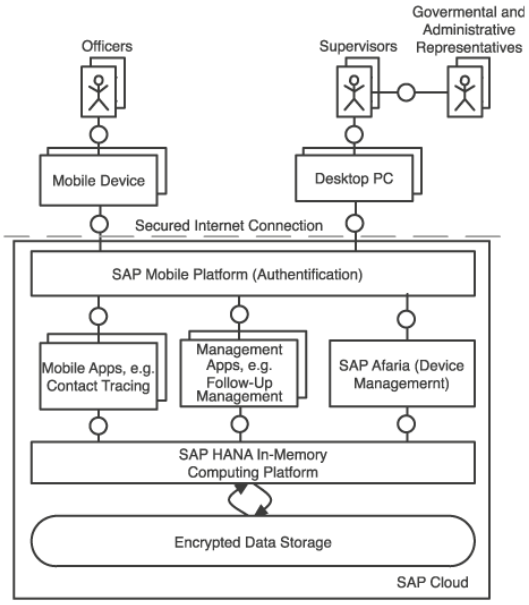
Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

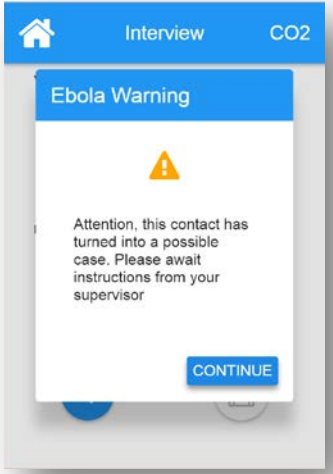

Contact Officer
Contact Officer
No
Yes
First Visit?
Interview Contact First Time
Interview Guide
Enter Demographic Data
Contact Tracing Form
Do Regular Check-Up
Contact Shows Symptoms
No
Yes
Potential New Contacts For This Case?
Enter New Contacts Or Contact Groups
Updated/New Contacts/Types Of Contact For Assignment
Record Contact Coordinates
Ask Contact About Direct Contacts Within Last 21 Days
Suspected Case Detected
Daily Report
Suspected Case Information With Contacts

SORMAS App for Contact Tracing by HPI: Software Architecture and Contact Tracing App

FIGURE 2
SORMAS software architecture



SORMAS: Surveillance and Outbreak Response Management System; VM: virtual machine.



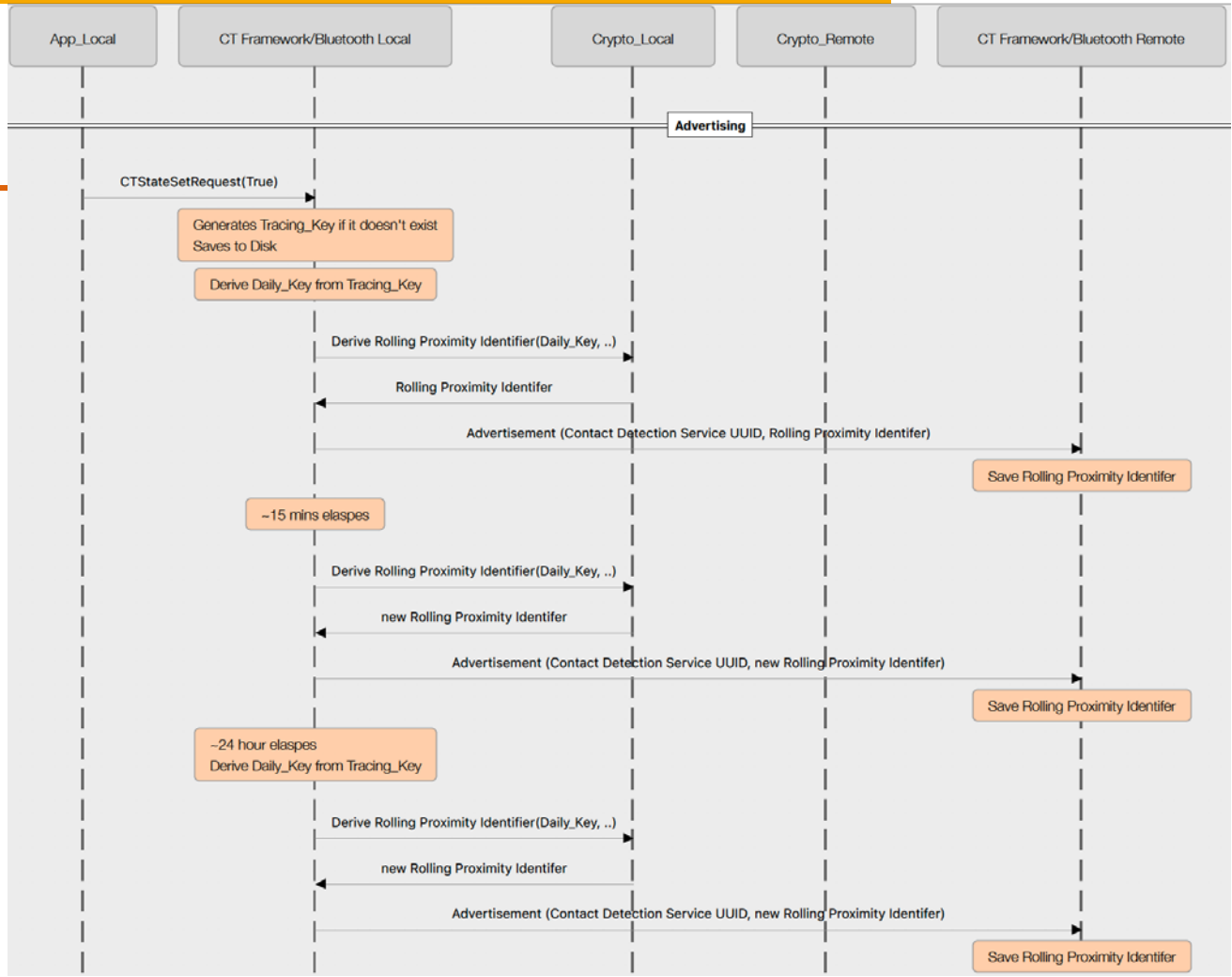
Fähnrich, C, Denecke, K, Adeoye, O O, Benzler, J, Claus, H, Kirchner, G, Mall, S, Richter, R, Schapranow, M-P, Schwarz, NG, Tom-Aba, D, Uflacker, M, Poggensee, G, Krause, G: Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa, Eurosurveillance, 20, 21071 (2015)

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

Contact Tracing 2020: GAEN

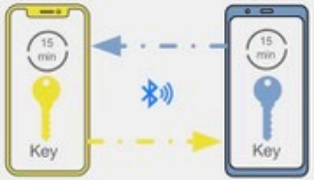
- Google/Apple Exposure Notification (GAEN) API implemented by OS vendors
- Privacy preserving, decentralized contact tracing
- Builds on near-field communication via Bluetooth
- GAEN implemented by 30+ countries apps (in 2022)



Contact Tracing 2020: GAEN

Alice and Bob meet each other for the first time and have a 10-minute conversation.

Their phones exchange anonymous identifier beacons (which change frequently).



A few days later...

Bob is positively diagnosed for COVID-19 and enters the test result in an app from a public health authority.

With Bob's consent, his phone uploads the last 14 days of keys for his broadcast beacons to the cloud.

Apps can only get more information via user consent

~14 day temporary store

Google

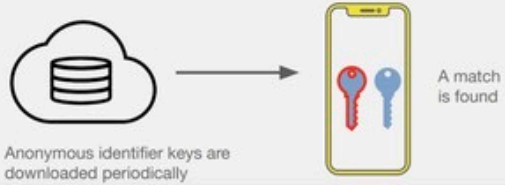
Contact Tracing 2020: GAEN

Alice continues her day unaware she had been near a potentially contagious person.

Alice sees a notification on her phone.

Sometime later...

Alice's phone periodically downloads the broadcast beacon keys of everyone who has tested positive for COVID-19 in her region. A match is found with the Bob's anonymous identifier beacons.



Alice's phone receives a notification with information about what to do next.

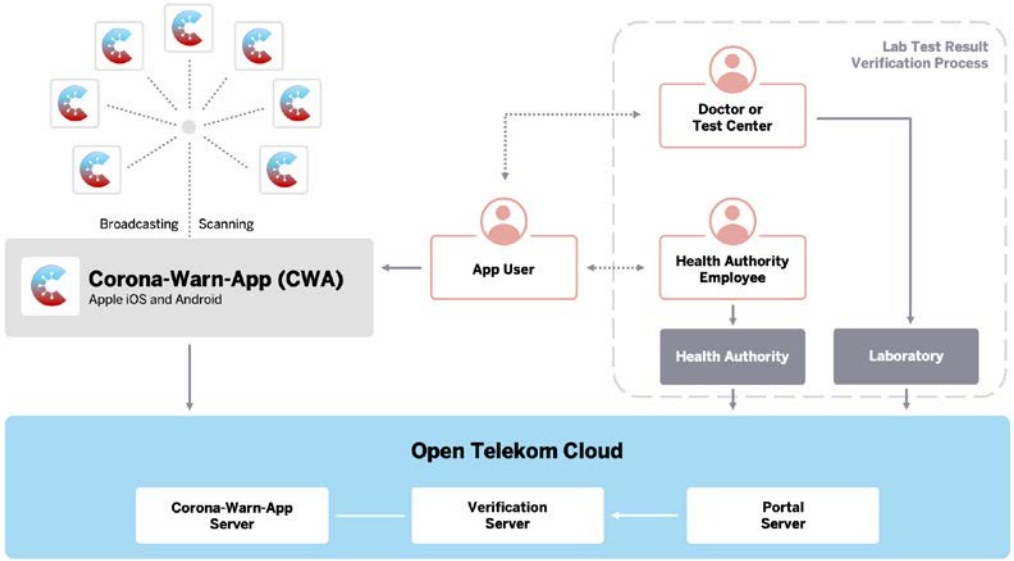
Google

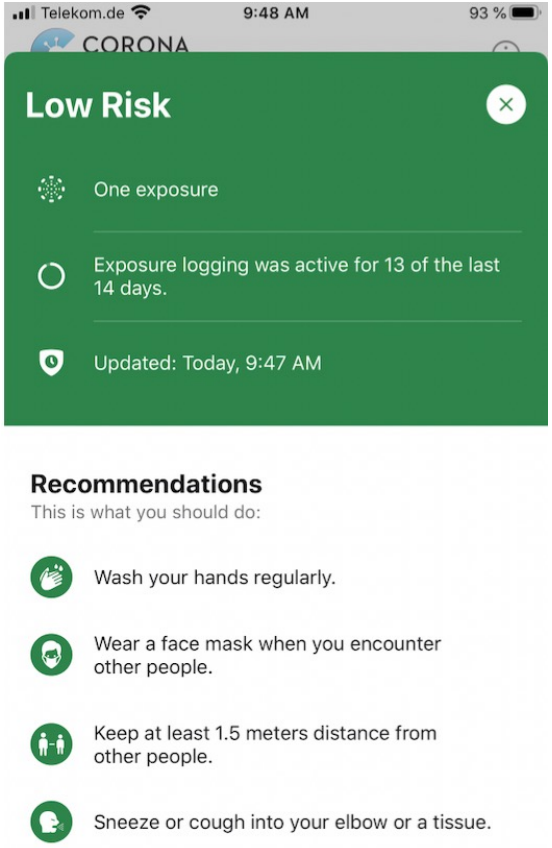
Corona Warn App:
A GAEN Frontend made in Germany

CORONA WARN-APP

Hasso Plattner Institut

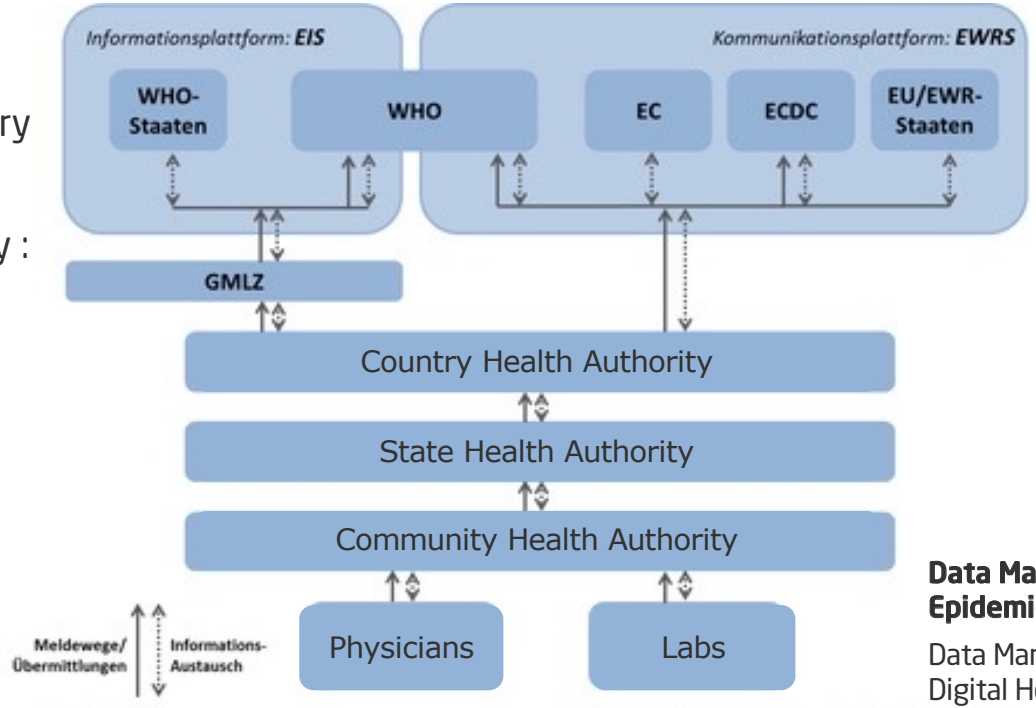
- Freedom to use
- Exchange of device IDs only
- No need for registration of personal data





Surveillance organized in EU (2020)

- Case report forms:
  - Prepared and sent by Laboratory Information System (LIS)
  - Send by fax to health authority :
    - Form is printed from LIS and hand-made notes are added
    - Form is filled by hand
    - Full medical report is send

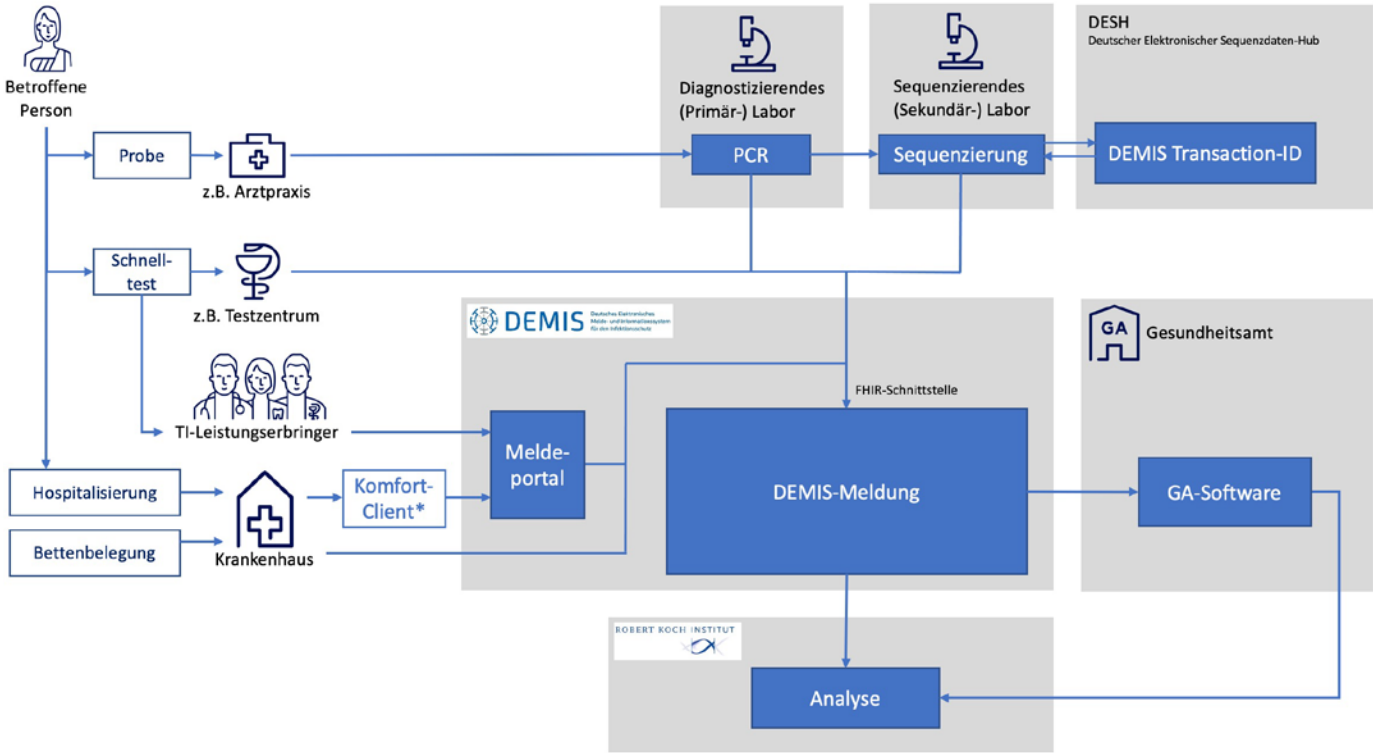


Surveillance in Germany (2024)

DEMIS Deutsches Elektronisches Melde- und Informationssystem für den Infektionsschutz

Hasso Plattner Institut

- Coordinated by RKI
- German Electronic Report and Information System for Infection Control
- Implemented as service of the German gematik
- Supported by Fraunhofer FOKUS
- Reports are exchange as FHIR/HL7 format



*: Der Komfort-Client (für das DEMIS Meldeportal) ist notwendig zur Authentifikation als Krankenhaus angeschlossen an die TI. Er ist kein DEMIS Produkt.

Virus Variant Surveillance: Variants under Monitoring (VUM)

- Genetic changes that are suspected to affect virus characteristics with some indication that it may pose a future risk
- Evidence of phenotypic or epidemiological impact is currently unclear
- Enhanced monitoring and repeat assessment pending new evidence is required

https://www.who.int/activities/tracking-SARS-CoV-2-variants

Virus Variant Surveillance: Variants of Interest (VOI)

- Genetic changes that are predicted / known to affect virus characteristics, e.g. transmissibility, disease severity, immune, diagnostic or therapeutic escape and
- Identified to cause significant
  - Community transmission or multiple clusters,
  - Present in multiple countries with increasing relative prevalence alongside increasing number of cases over time, or
  - Other apparent epidemiological impacts to suggest an emerging risk to global public health
- Past COVID-19 examples: epsilon, zeta, eta, theta, iota, kappa, lambda, mu

https://www.who.int/activities/tracking-SARS-CoV-2-variants

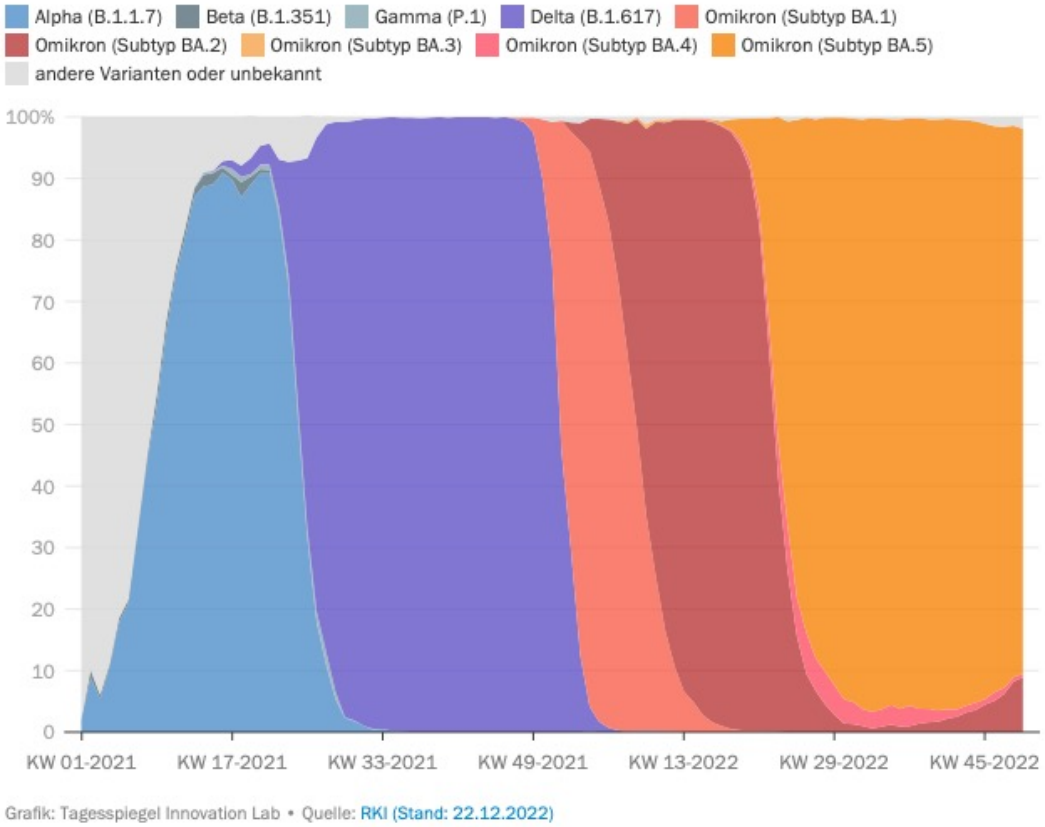
Virus Variant Surveillance: Variants of Concern (VOC)

- Increase in transmissibility or detrimental change in COVID-19 epidemiology, or
- Increase in virulence or change in clinical disease presentation, or
- Decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics.

- COVID-19 examples
  - VOC as of Jan 2023: Omicron (B.1.1.259), VUM (Nov 24), VOC (Nov 26, 2021)
  - Past VOCs: alpha, beta, gamma, delta

https://www.who.int/activities/tracking-SARS-CoV-2-variants

SARS-CoV-2 Variants over Time Germany (Jan 2021– Dec 2022)



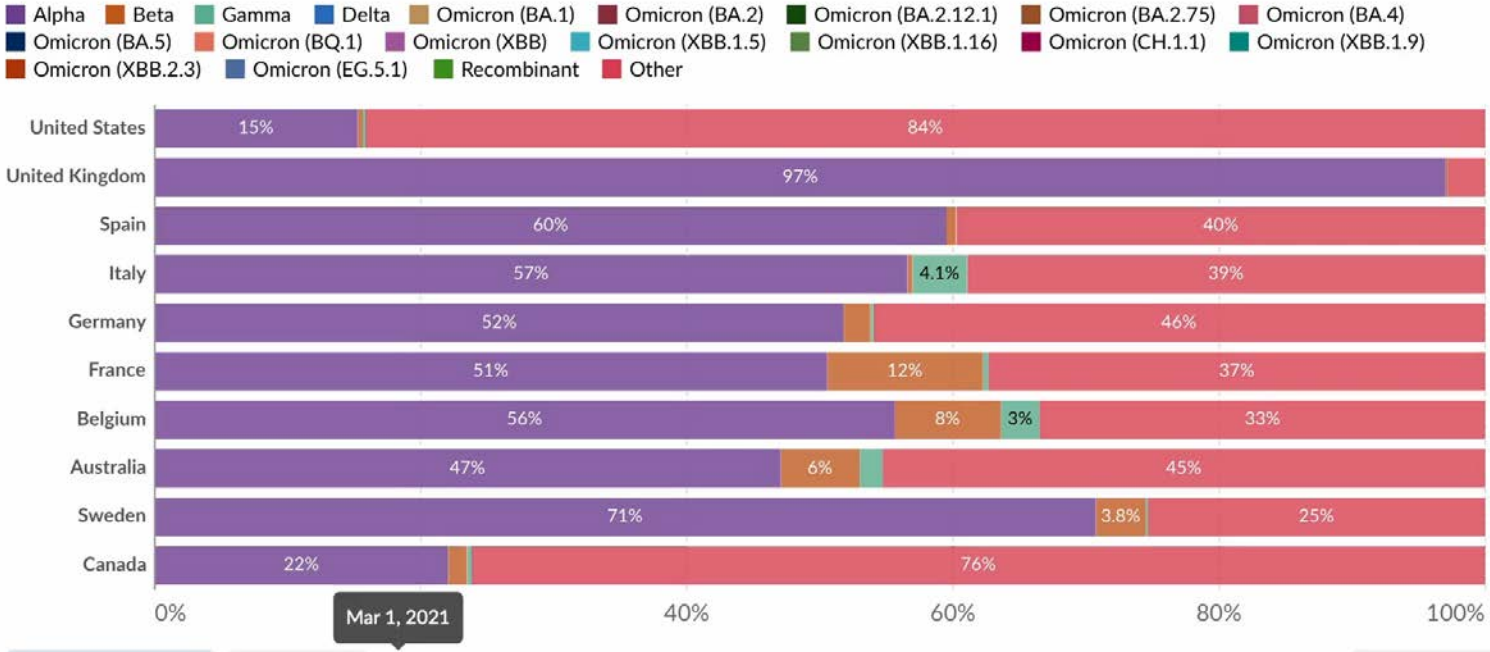
Grafik: Tagesspiegel Innovation Lab • Quelle: RKI (Stand: 22.12.2022)

SARS-CoV-2 sequences by variant, Mar 1, 2021

Our World in Data

The share of analyzed sequences in the preceding two weeks that correspond to each variant group. This share may not reflect the complete breakdown of cases since only a fraction of all cases are sequenced.

Table | Chart | Edit countries and regions | Settings



Play time-lapse | Mar 1, 2021 | Dec 18, 2023

Data source: GISAID, via CoVariants.org – Last updated 18 December 2023 – Learn more about this data

Note: Recently-discovered or actively-monitored variants may be overrepresented, as suspected cases of these variants are likely to be sequenced preferentially or faster than other cases.

OurWorldInData.org/coronavirus | CC BY

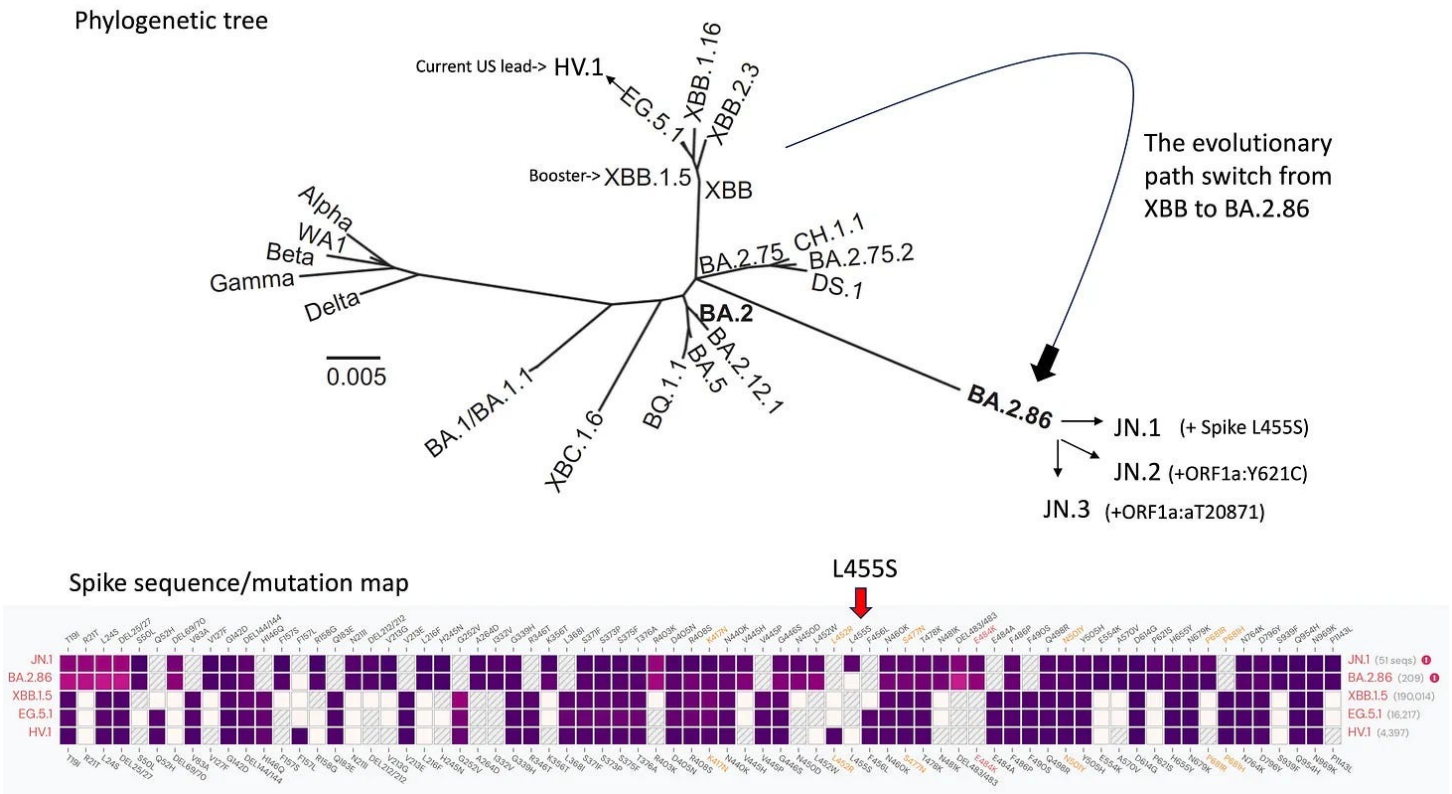
https://ourworldindata.org/grapher/covid-variants-bar

Hasso Plattner Institut

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

Latest SARS-CoV-2 Subtypes
From XBB to BA.2.86



https://erictopol.substack.com/p/from-a-detour-to-global-dominance

Data Management for Epidemiology

Data Management for Digital Health, Winter 2023

What To Take Home?

- Epidemiological modeling can help to estimate disease spread
- Epidemiological surveillance is key for management of countermeasures to contain infectious diseases
- Visualization helps to understand numbers and might help to discover cause-effect relationships
- Contact tracing and quarantine help to reduce spreading speed and to provide reaction time