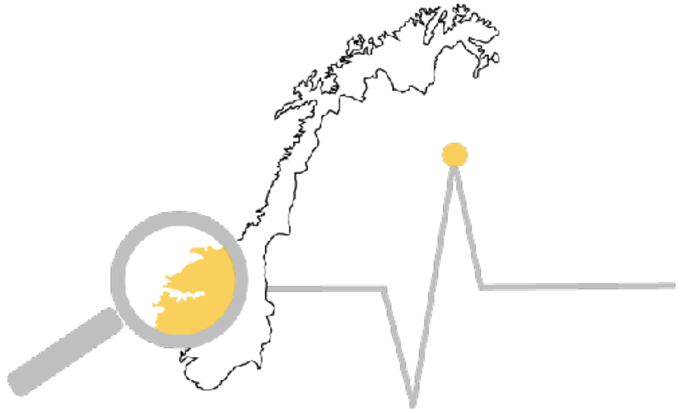


Public Health Surveillance and Reporting

An Open Source Approach

Chi Zhang

@Andreasheenn chi.zhang@medisin.uio.no

OCBE (Oslo Center for Biostatistics and Epidemiology) @ UiO

ex-FHI (Norwegian Institute of Public Health)

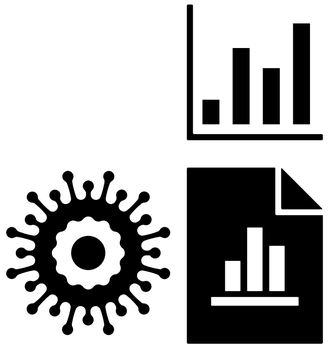
2023-03-30 Open Science Lunch @ UiO

Github: `sykdomspulsen-org, csids`

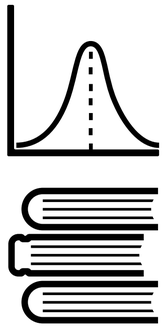
Website: https://docs.sykdomspulsen.no, https://www.csids.no

About me

FHI Researcher (statistician + R developer)
Sykdomspulsen team, infectious diseases



UiO (Faculty of Medicine)
Researcher / lecturer in biostatistics



2020 2021 2022 2023

About this talk

Public health surveillance and reporting

Introduction
Surveillance vs Research

Open PH data and tools

Motivation
Open data, open source software

Sykdomspulsen / CSIDS

Open-source, automated platform for surveillance, analysis and reporting

Challenges and ways forward

“How to prevent the next pandemic”

Disclaimers

The opinions are my own, and do not reflect the views of my employer.

Regulations and policies are constantly changing. Please check the official documents for each data source for most updated information.

PH surveillace and reporting in a nutshell 

Public health surveillance

Introduction

Continuous, systematic collection, analysis and interpretation of **health-related** data (WHO definition)

Routine (e.g. daily, yearly);

different levels of geo-locations;

Some are event related

Surveillance cycle

Collection,
Analysis,
Dissemination,
Action

Infectious diseases
(influenza, covid ⋯),

Mortality, causes of death,

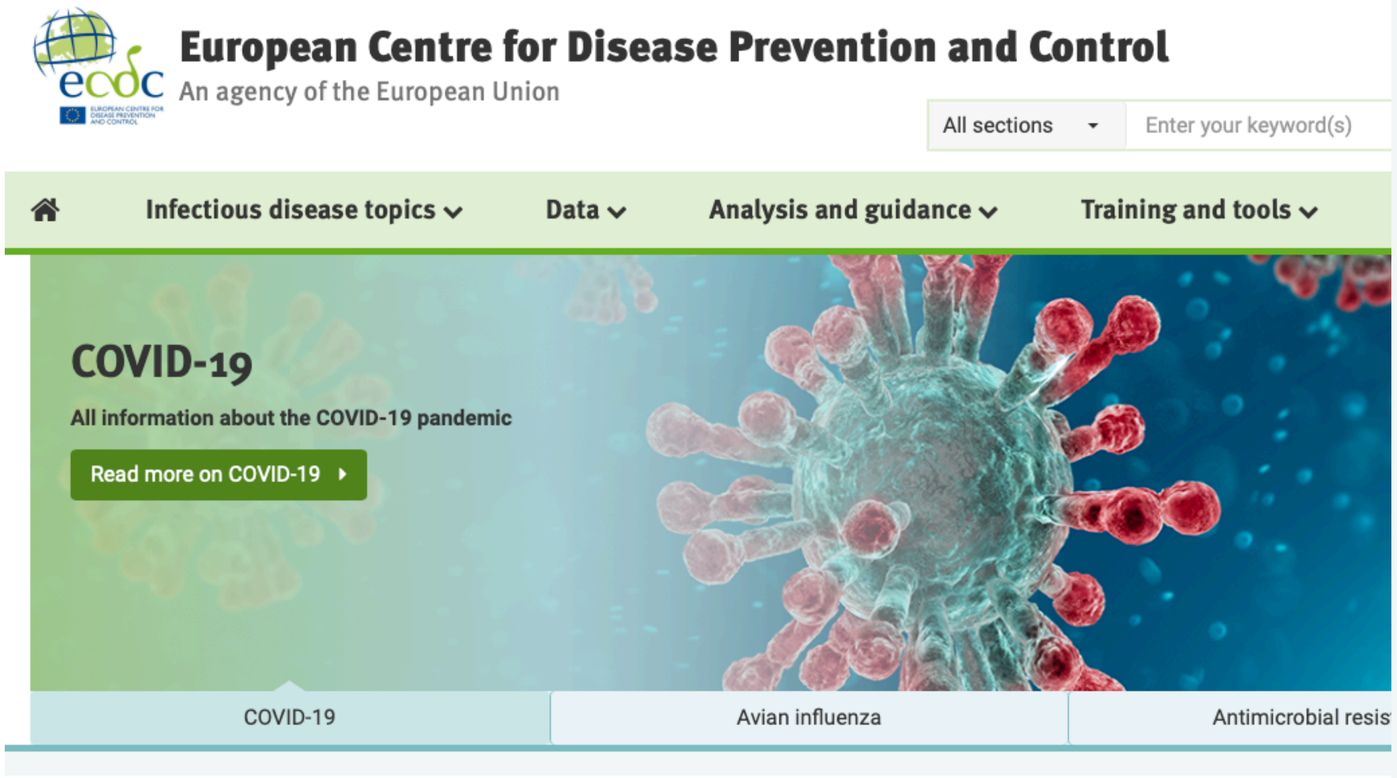
Chronic diseases
(cancer, diabetes ⋯),

Meteorological and natural disasters
(heatwave, hurricane),

Others
(firearm injury in US ⋯)

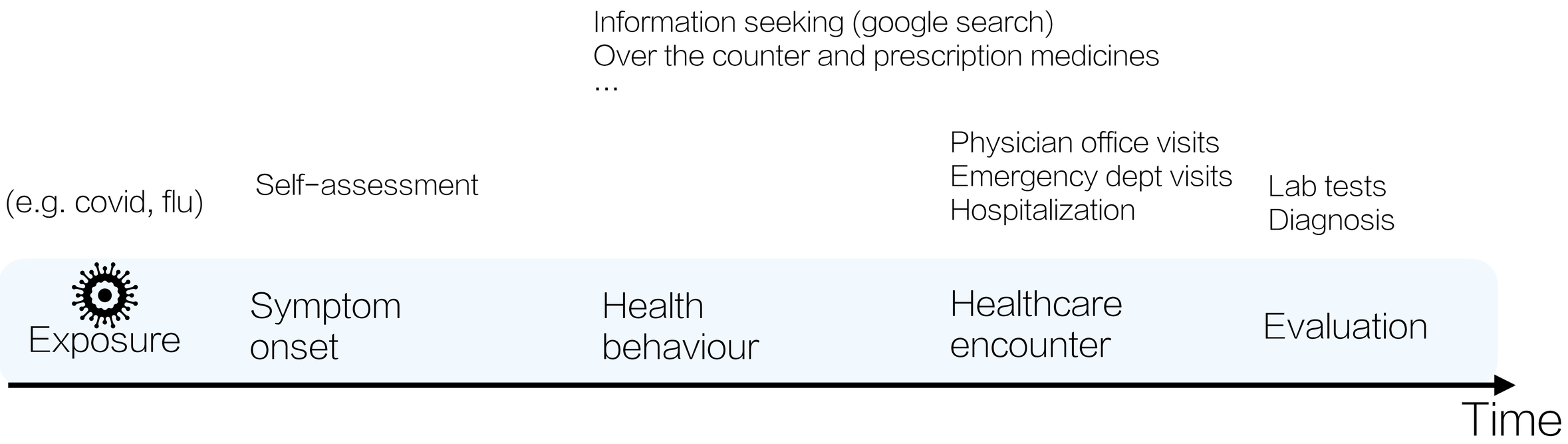
Public health surveillance

Introduction



Surveillance on infectious diseases

Where does data come from?



Data collected by **healthcare professionals**, then some are sent to **regional / national PH authority** for actions

Inspiration: Coursera JHU course on surveillance

Surveillance on infectious diseases

Compared to “research”

Surveillance cycle: CADA

Tasks	Research	Surveillance (especially disease surveillance with **outbreak** potential)	Challenges in surveillance
Collection	Fixed datasets	New data, multiple sources, every day	Data aggregation and cleaning
Analysis	Various	Descriptive (e.g. count, trend)	Flexibility age, sex, time, location groups
Dissemination	(Journal) publication	Reports, graphs, numbers, datasets	Scale and consistency
Action		Policy responses (effect immediately)	Rapid new task development

Surveillance methods can NOT live without research;

Timely **reporting** is a core task in public heatlh surveillance

Inspiration: Coursera JHU course on surveillance

Open PH data and tools

Where to find data and reports?

Public health agency

Collect, process (e.g. censoring), analyse data

Internal reports at various frequencies

Release to the public some data and reports

Used by:

Media
(e.g. Aftenposten, VG)

International networks
(e.g. WHO, ECDC, EuroMOMO)

Open public health

A mindset

"Public" health data, open access everything?

It is **impractical** and **unnecessary** to open ALL the real-time public health data

- privacy protection, censoring, aggregation
- delays, validation
- technical challenges (e.g. data pipeline, maintenance)
- Misinterpretation could cause panic
- ...

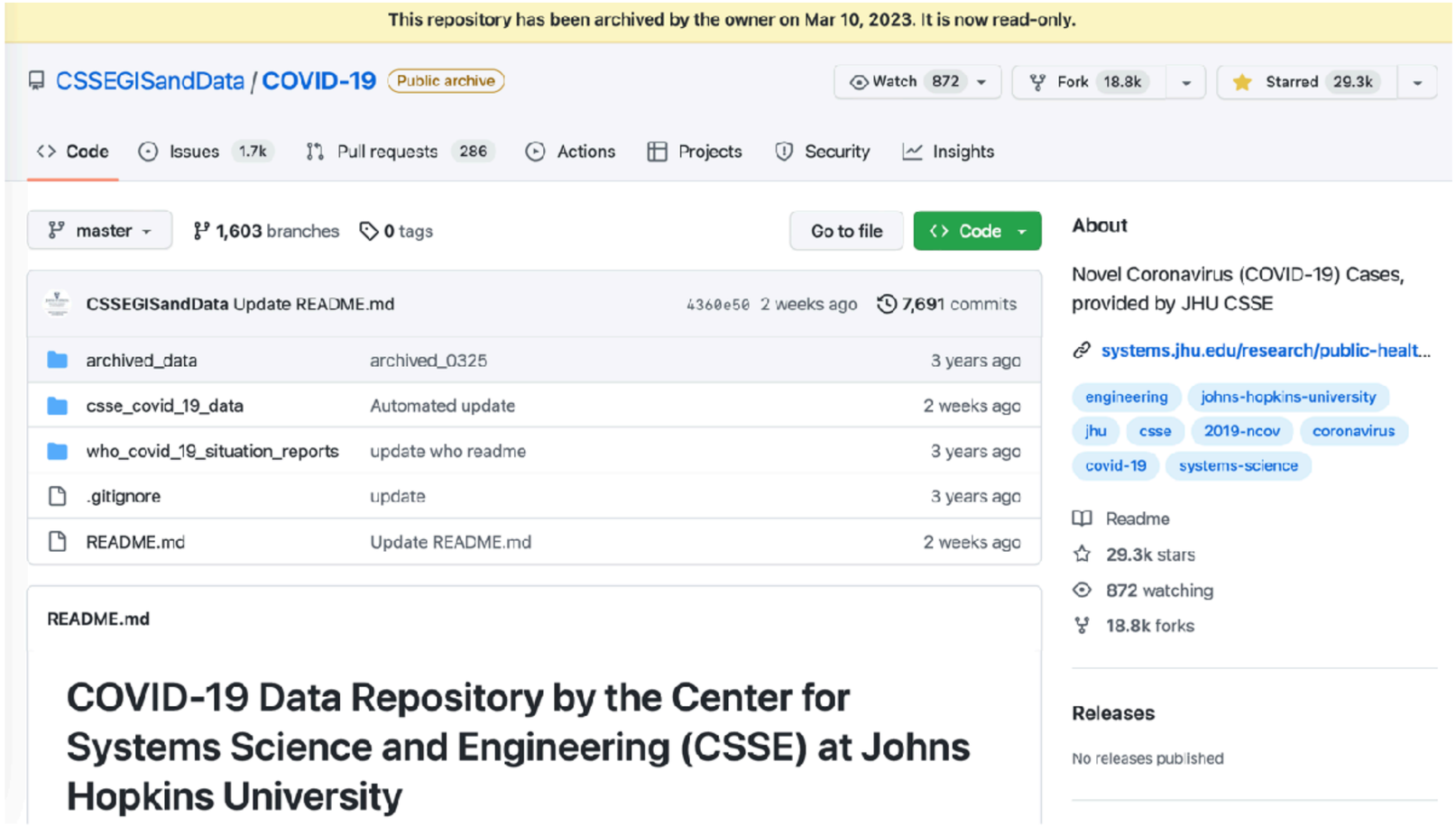
The existing public data can be used in a **more collaborative, and accessible** way to

- build trust and transparency
- national and international collaboration
- facilitate research
- ...

Open source tools and software can be very useful in this process.

Public data on GitHub

Johns Hopkins University COVID repository



Covid cases, deaths, vaccines etc

National, county/state level

Widely used for visualisation

Archived 2023.3.10

29.3k stars, 18.8k forks

Public data on GitHub

Our World in Data

Research and data

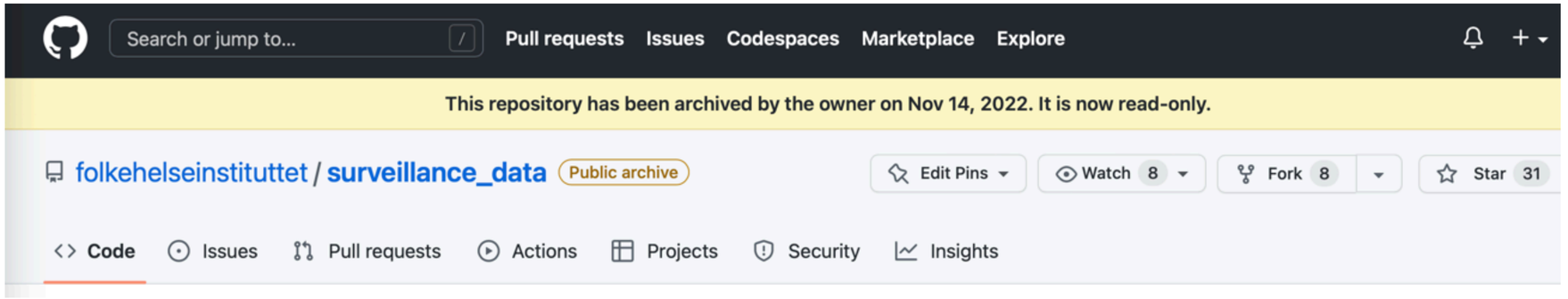
All free, open access and open source

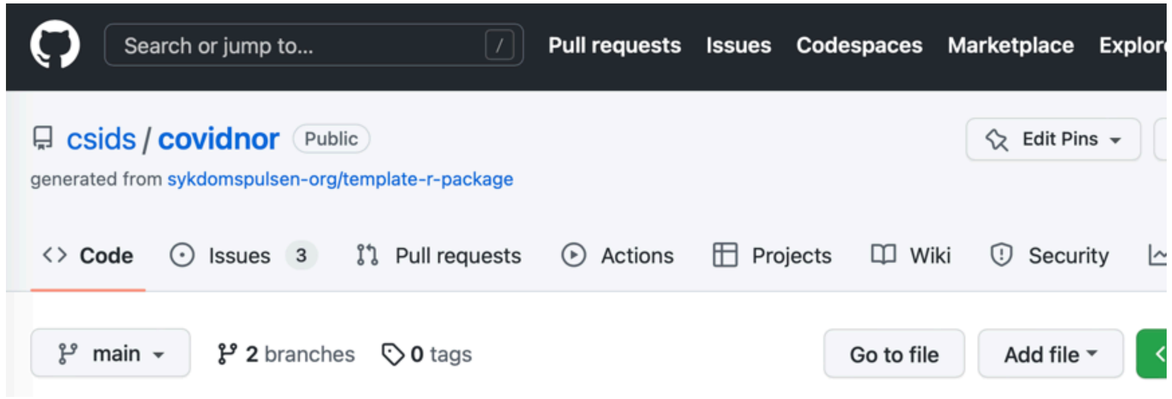
Public data on GitHub

Our World in Data

Public data on GitHub

Covid data in Norway





Daily data on cases, hospitalisation, vaccine etc

By age groups, sex, location

Used by VG, archived 2022.11.14

Processed covid data in Norway in **new repository**
(work in progress)

Open source software

Should I switch?

Consider the following aspects:

Cost

Ease of use (e.g well documented)

Efficiency (e.g. automation)

Collaboration and teamwork

Reproducibility

Research and new method adoption

VS

Surveillance x Open source tools

= Sykdomspulsen

Sykdomspulsen

The Disease Pulse

A team of 8* members at **Norwegian Institute of Public Health**
Started in 2012. Key role in Covid-19 pandemic

Sykdomspulsen Core and R packages were migrated to **CSIDS project** in November 2022.

Automated real time public health surveillance platform

Daily** data input

Daily analysis

Daily delivery

Infectious diseases

- Covid19
- Influenza
- Other respiratory
- Gastrointestinal
- Other infections

Mortality

- Excess mortality
- Cause of death

Data cleaning and censoring

Statistical analysis

Graph making and reporting

* 2021-22
** Mortality surveillance is either weekly or yearly

Sykdomspulsen

Automated situational reports

Situational reports on covid cases, hospitalisation, vaccine etc

1 nation + 11 counties + 356 municipalities

Everyday, before **7am**

Used by FHI leadership and Ministry of Health

- Agder_fylke_dagsrapport_covid19_2021-10-20.docx
- Innlandet_fylke_dagsrapport_covid19_2021-10-20.docx
- Møre_og_Romsdal_fylke_dagsrapport_covid19_2021-10-20.docx
- Nordland_fylke_dagsrapport_covid19_2021-10-20.docx
- Oslo_fylke_dagsrapport_covid19_2021-10-20.docx
- Rogaland_fylke_dagsrapport_covid19_2021-10-20.docx
- Troms_og_Finnmark_fylke_dagsrapport_covid19_2021-10-20.docx
- Trøndelag_fylke_dagsrapport_covid19_2021-10-20.docx
- Vestfold_og_Telemark_fylke_dagsrapport_covid19_2021-10-20.docx
- Vestland_fylke_dagsrapport_covid19_2021-10-20.docx
- Viken_fylke_dagsrapport_covid19_2021-10-20.docx

COVID-19

Dagsrapport – nasjonalt og per fylke

onsdag 20. oktober 2021

Demo

Dagens foreløpige (u.off) tall. Rapporten er generert kl. . Der annet ikke er oppgitt er figurene basert på prøvedato i MSIS. Tallene er midlertidige og kan bli endret.

Status oppdatering

Totalt nye tilfeller ble registrert siste døgn. De to foregående dagene ble det registrert henholdsvis tilfeller.

Til sammenligning ble det for en uke siden, den 06.10.2021 rapportert registrerte tilfeller siste døgn.

Antall meldte basert på prøvedato så langt uke 41 er mot på samme tid sist uke (uke 40). Antall meldte basert på registrert dato så langt denne uka (uke 41) er , mot på samme tid sist uke (40).

Totalt er personer vaksinert mot covid-19 i Norge, av disse er personer vaksinert med både 1 og 2. dose med koronavaksine av hele befolkningen er vaksinert med minst en dose og vaksinert med to doser med koronavaksine. Blant personer 18 år og eldre er % vaksinert med minst en dose med koronavaksine, og av disse er % vaksinert med to doser med koronavaksine. Blant personer 45 år og eldre er % vaksinert med minst en dose, og % av personer 65 år og eldre er vaksinert med minst en dose med koronavaksine. Data er hentet fra BeredtC19, SYSVAK, per 12.10.2021.

Sykdomspulsen

Automated situational reports

35 pages, 17 tables, 21 figures

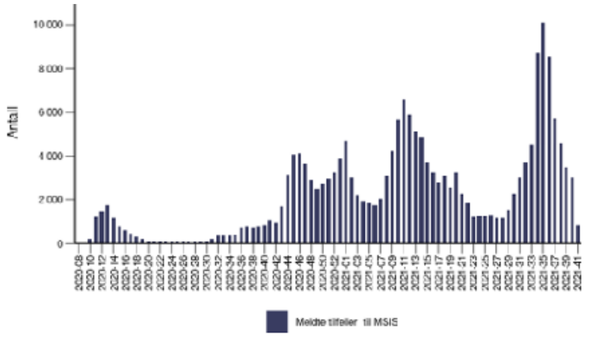
Tabell 1 Covid-19 status og utvikling, uke 2021-37 til uke 2021-41.

Demo

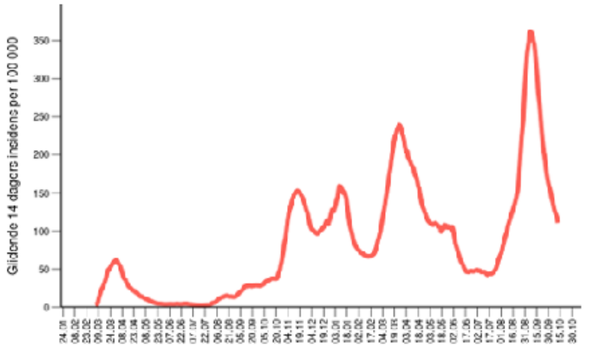
Overvåkingssystem/ Indikatorer	2021-37	2021-38	2021-39	2021-40	2021-41	t av de siste 5 ukene
Meldte tilfeller til MSIS (prøvedato)						
Meldte tilfeller til MSIS (registrertdato)						
Antall personer testet for SARS-CoV-2 (PCR)						
Nye covid-19 positive pasienter innlagt i sykehus (alle årsaker)						
Nye pasienter innlagt i sykehus med covid-19 som hoved-årsak						
Nye pasienter med bekreftet covid-19 innlagt i intensiv-avdeling						
Covid-19-assosierte dødsfall						

*Visualiseringen på nettsidene våre oppdateres ca kl 13.00 hver dag.

Merk: Denne tabellen og visualiseringen på www.fhi.no viser antall personer testet med PCR, vi jobber med å inkludere antigen hurtigtester. Se ukesrapport for framstilling av antall testede med PCR og hurtigtester samlet. Totalt har [redacted] ersoner blitt diagnostisert i Norge frem til kl 24.00, 12.10.2021, og tilsammen [redacted] personer er registrert testet for covid-19 (per 12.10.2021).



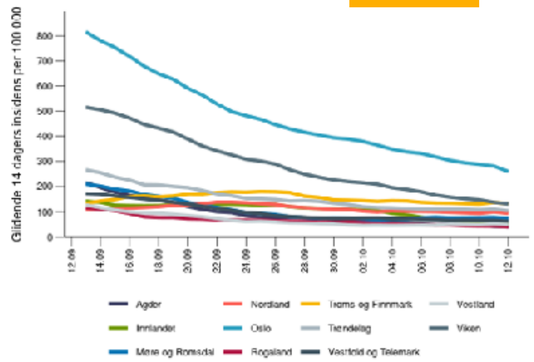
Figur 1. Antall tilfeller av covid-19 per uke basert på prøvedato gjennom hele pandemiperioden, Norge.



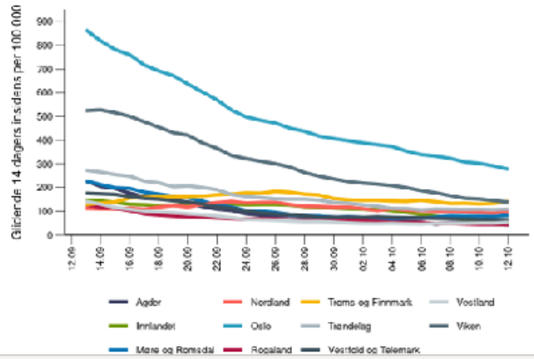
Figur 2. Glidende 14-dagers insidens per 100 000 innbyggere per dag basert på prøvedato gjennom hele pandemiperioden, Norge.

Fylker

Demo



Figur 5. Antall tilfeller per 100 000 innbyggere per 14 dager etter fylke og prøvedato, uke 2021 37 til uke 2021 41.



Sykdomspulsen

Automation saves time and cost

Historically (pre 2020.12) made **manually**, **in the early morning**, for **selected locations**

Overtime estimated by one employee work from 6 to 8am for one year (2h per day)
One employee is allowed 200h overtime per year

National + 11 counties + 356 municipalities = 368 reports
From Secure zone (data extraction) to reports = **40**min (8 CPU in parallel)

One year deliverables	Manual report	Sykdomspulsen
Number of reports	5 * 365 = 1 825	368 * 365 = 134 320
Overtime hours	700	0
Number of people needed	4	0
Overtime costs	700 000 kr	0

700 000 kr is approximately 67 400 Euro, 70 245 USD (2022.06.15)

From data to report

Secure zone (like TSD)

Ordinary zone

Data → Anonymise → NHN sluse (export) → Process → Analyse → Deliver

MSIS
MSIS mortality
MSIS lab
NIR/NOPAR
SYSVAK

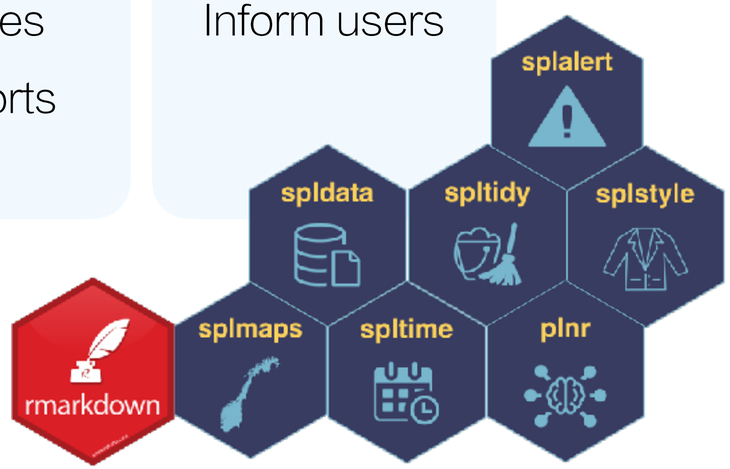
By Sykdomspulsen
By lab department
By team overvåking

data.rds

Clean data
Censor data
Check updates
Send alerts

Statistics
Graphs
Tables
Reports

Manual checks
Send reports
Inform users



MSIS: Norwegian Surveillance System for Communicable Diseases
NIR: Norwegian Intensive Care Registry
SYSVAK: Norwegian Immunisation Registry
NOPAR: Norwegian Pandemic Registry

From data to report

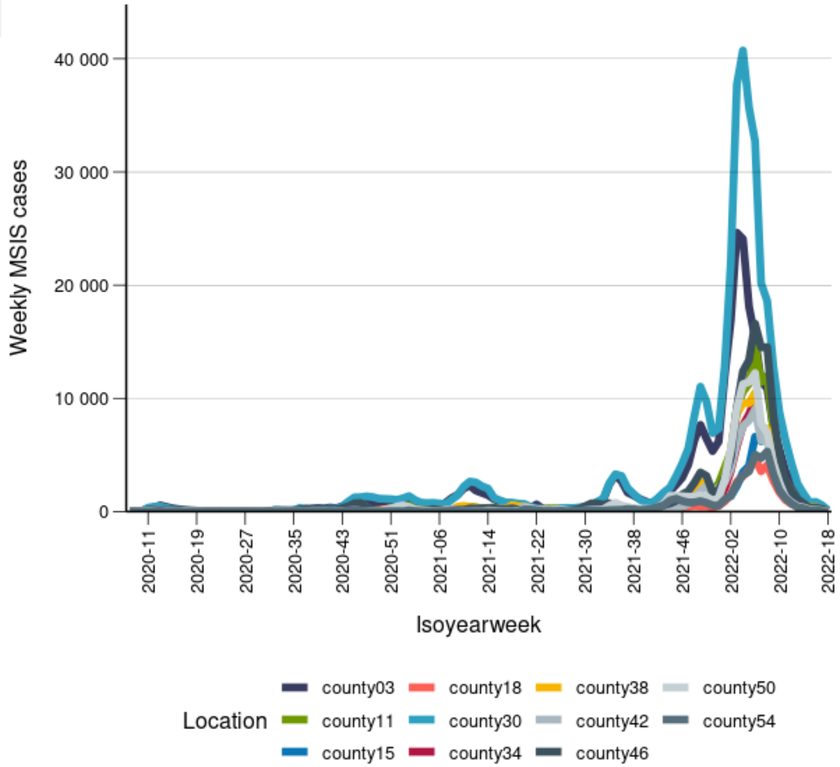
Example: daily report for covid trend

Complete example please check vignette documentation in each package

```
d_msis <- spltidy::covid19_msis_cases_by_time_location
# check the column names
```

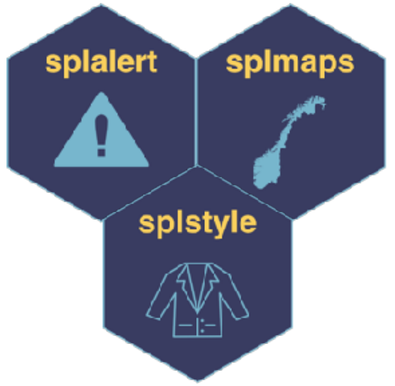
	[unified] isoyearweek	[unified] season	[unified] seasonweek	[unified] calyear	[unified] calmonth
	<character>	<character>	<numeric>	<integer>	<integer>
	NA=0 %	NA=0 %	NA=0 %	NA=13 %	NA=13 %
1:	2020-08	2019/2020	31	2020	2
2:	2020-08	2019/2020	31	2020	2
3:	2020-08	2019/2020	31	2020	2
4:	2020-09	2019/2020	32	2020	2
5:	2020-09	2019/2020	32	2020	2
...					
11024:	2022-14	2021/2022	37	NA	NA
11025:	2022-15	2021/2022	38	NA	NA
11026:	2022-16	2021/2022	39	NA	NA
11027:	2022-17	2021/2022	40	NA	NA
11028:	2022-18	2021/2022	41	NA	NA

	[unified] calyearmonth	[unified] date	[context] covid19_cases_testdate_n	[context] covid19_cases_testdate_pr100000
	<character>	<Date>	<integer>	<numeric>
	NA=13 %	NA=0 %	NA=0 %	NA=0 %
1:	2020-M02	2020-02-21	0	0
2:	2020-M02	2020-02-22	0	0
3:	2020-M02	2020-02-23	0	0
4:	2020-M02	2020-02-24	0	0
5:	2020-M02	2020-02-25	0	0
...				
11024:	NA	2022-04-10	6888	126.9614
11025:	NA	2022-04-17	3635	67.0013
11026:	NA	2022-04-24	3764	69.379
11027:	NA	2022-05-01	2243	41.3436
11028:	NA	2022-05-08	502	9.253



From data to report

Example: daily report for covid trend



splalert 2022.6.15 Get started Reference Articles Changelog

Determine the short term trend

Source: R/short_term_trend.R

Determine the short term trend

Usage

```
short_term_trend(
  x,
  numerator,
  denominator = NULL,
  prX = 100,
  trend_days = 42,
  remove_last_days = 0,
  forecast_days = trend_days,
  trend_isoweeks = ceiling(trend_days/7),
  remove_last_isoweeks = ceiling(remove_las
  forecast_isoweeks = trend_isoweeks,
  numerator_naming_prefix = "from_numerator
  denominator_naming_prefix = "from_denomin
  statistics_naming_prefix = "universal",
  remove_training_data = FALSE
)
```

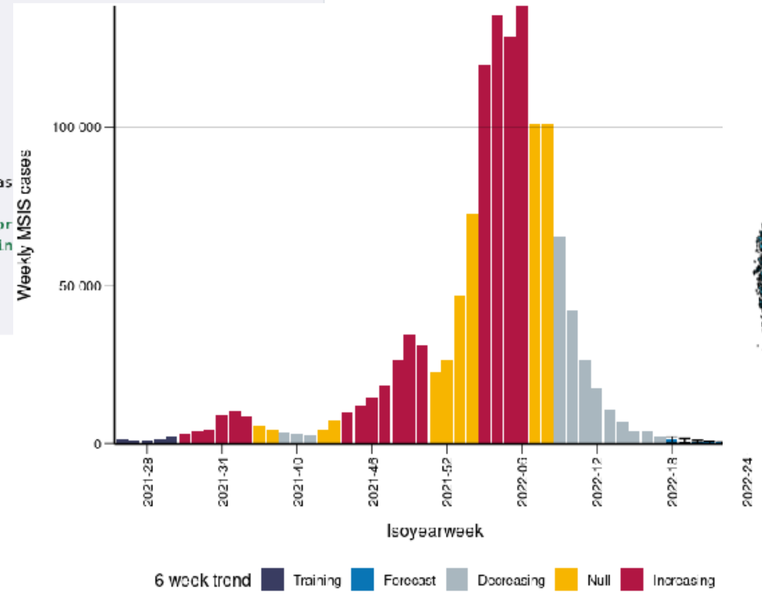
Arguments

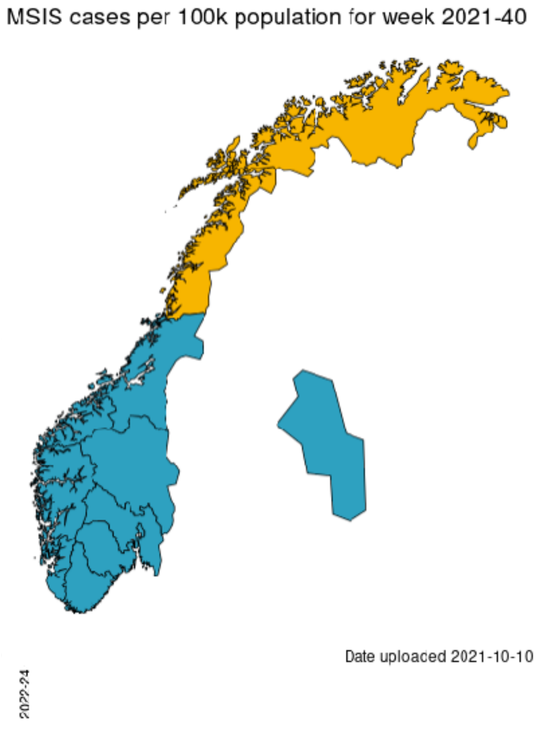
x
Data object

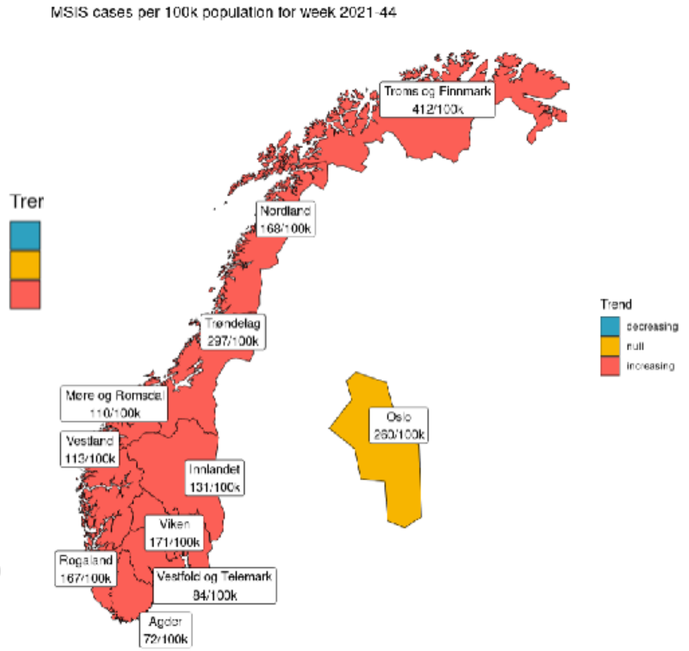
numerator
Character of name of numerator

denominator
Character of name of denominator (optional)

prX
If using denominator, what scaling factor should be used for numerator/denominator?







From data to report

Example: daily report for covid trend

```
---
title: "Test Report"
output:
  word_document:
    reference_docx: style_demo.docx
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Introduction and disclaimers

This is a demonstration for how to make automated reports with Rmarkdown ar
Word.

The dates, numbers, graphs and tables are for educational purposes only. Th
is made to teach R, not to be used in other circumstances.

The data used in this document is aggregated data, and it is public availab

This is how we make a report

A report has text. You're reading the text. **Standard markdown style works
well**.

We insert the figure made before.

![MSIS Covid-19 cases trend for week 2021-44](map_covid_demo.png)
```

COVID-19

Auto Report (demo for tutorial)

tirsdag 14. juni 2022

Introduction and disclaimers

This is a demonstration for how to make automated reports with Rmarkdown and MS Word.

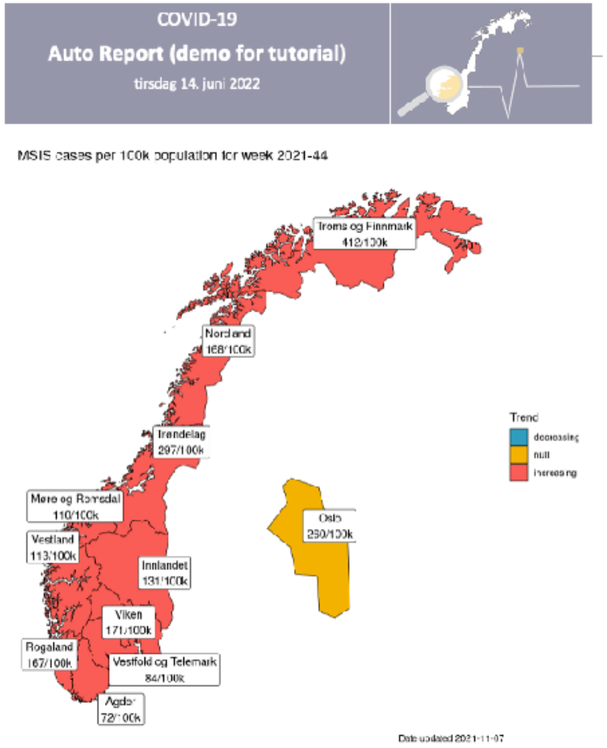
The dates, numbers, graphs and tables are for educational purposes only. The text is made to teach R, not to be used in other circumstances.

The data used in this document is aggregated data, and it is public available.

This is how we make a report

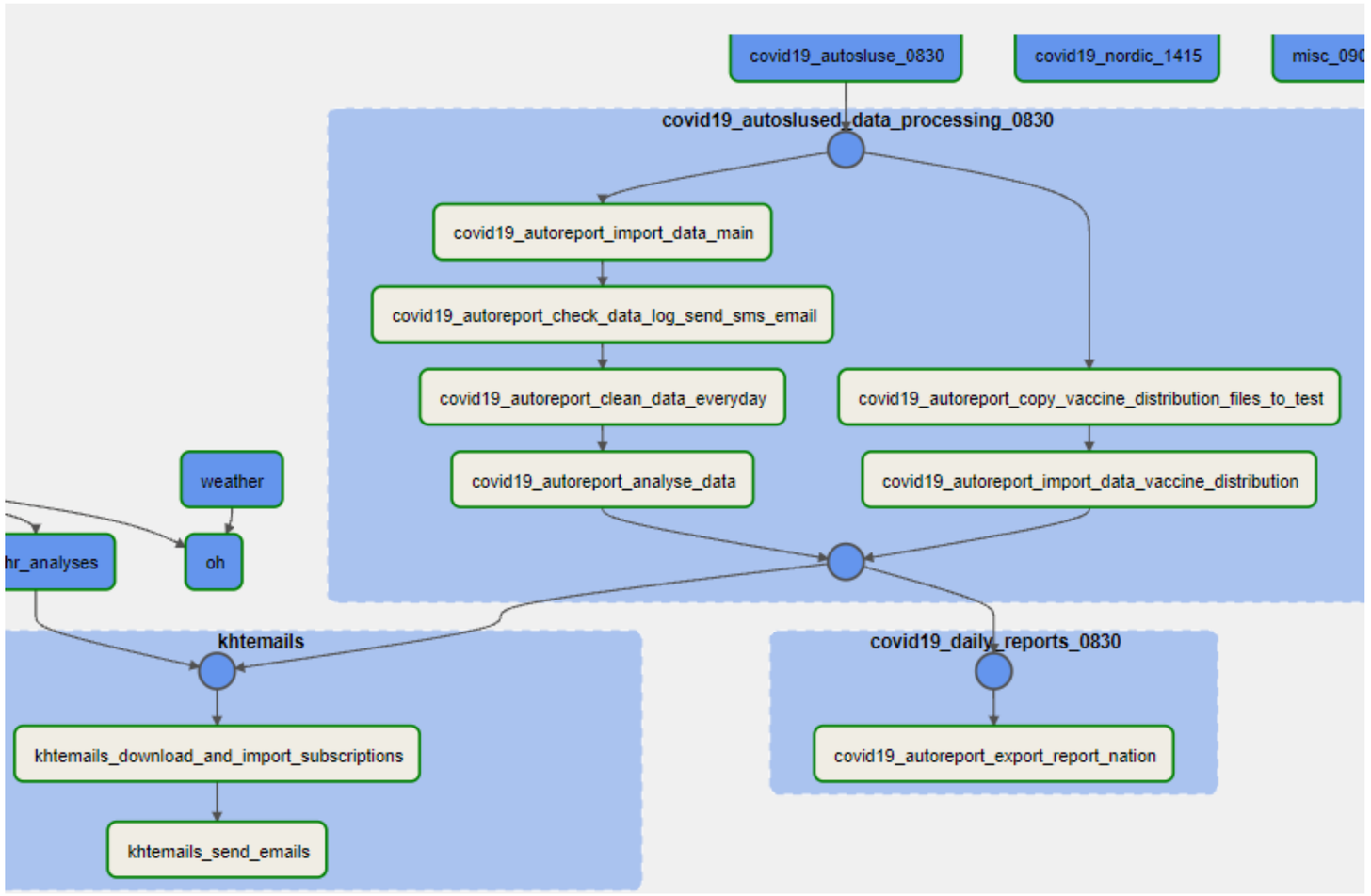
A report has text. You're reading the text. **Standard markdown style works here as well**.

We insert the figure made before.



Development and Automation

Airflow



We run 150+ tasks in airflow

Alert system: Email + SMS

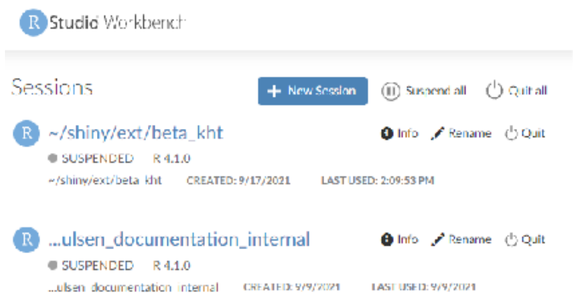
Quality assurance: vakt (debug, check report output)

Development and Automation

RStudio, Github, Docker,CI/CD

Rstudio workbench

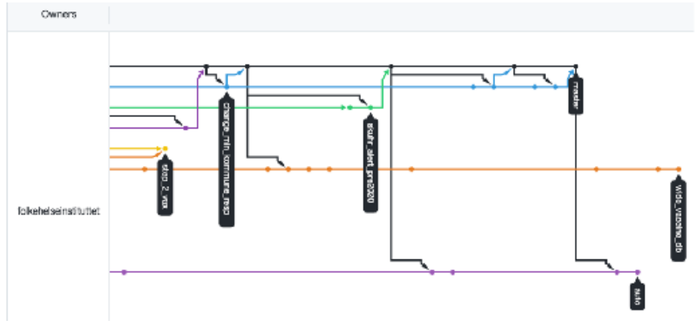
(Formerly Rstudio Server Pro)



100 USD per person per month

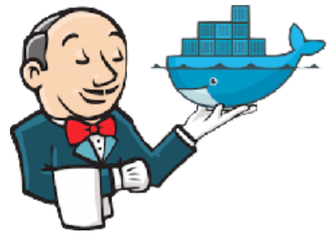
Github

One edit per branch



GoCD

Build docker image *



Contact our engineer for more details!

Risks, Challenges & Ways forward

This pandemic is over (?)

What about the next one

DAGENS Medisin

DM Debatt DM Pharma DM Arena DM Jobb Logg inn

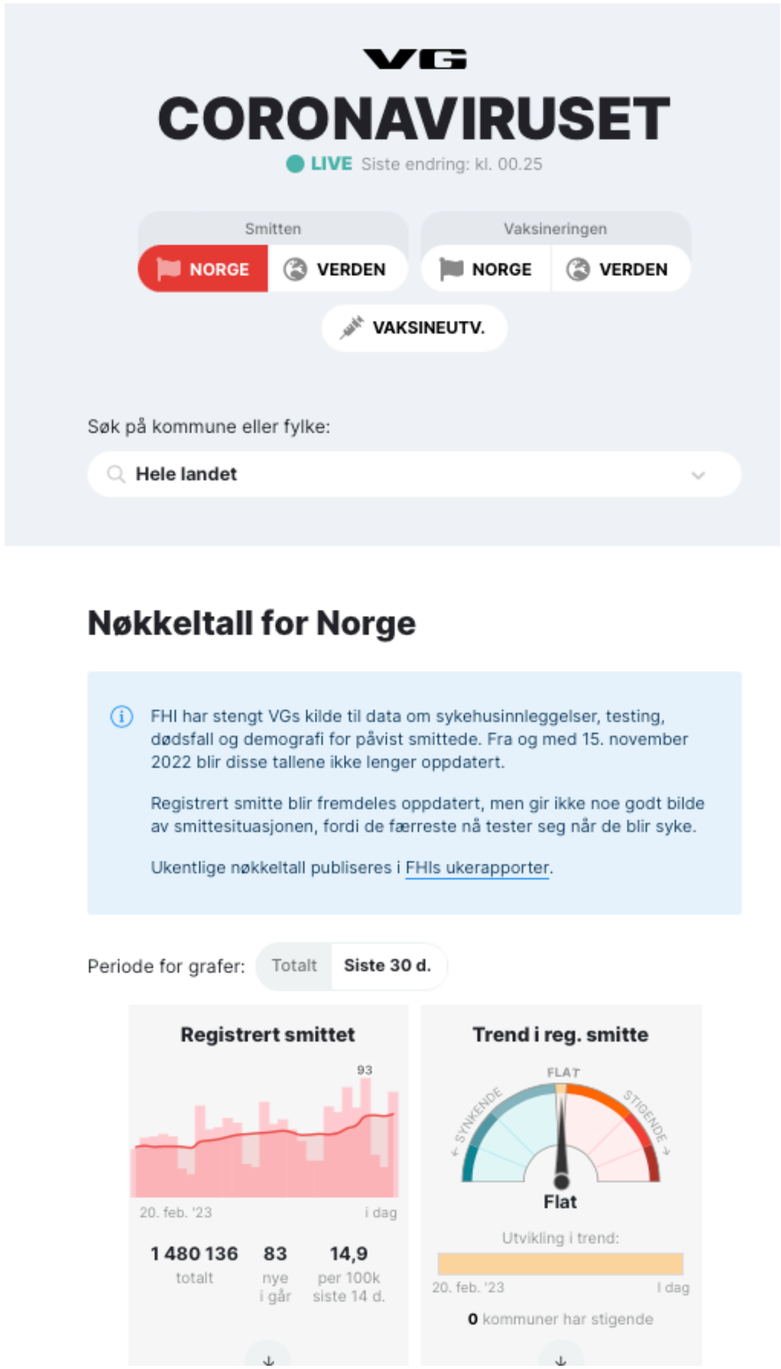
FHIs Sykdomspulsen for kommunehelsetjenesten legges ned på grunn av budsjettkutt

Nettsidene der blant andre kommuneleger, smittevernleger og statsforvaltere kan holde seg oppdatert på smittsomme sykdommer i sitt område, legges ned tirsdag.

Leni Aurora Brækhus

PUBLISERT Mandag 14. november 2022 - 19:09

Nettsiden for oppdaterte, daglige tall om covid-19, influensa, andre luftveisinfeksjoner, mage-tarminfeksjoner, vaksinasjonsdata for covid-19, vaksinasjonsdata for influensa og dødelighet legges ned fra og med 15. november, skriver FHI i en melding på sine nettsider.



VG

CORONAVIRUSET

LIVE Siste endring: kl. 00.25

Smitten: NORGE VERDEN

Vaksineringen: NORGE VERDEN

VAKSINEUTV.

Søk på kommune eller fylke:

Hele landet

Nøkkeltall for Norge

FHI har stengt VGs kilde til data om sykehusinnleggelser, testing, dødsfall og demografi for påvist smittede. Fra og med 15. november 2022 blir disse tallene ikke lenger oppdatert.

Registrert smitte blir fremdeles oppdatert, men gir ikke noe godt bilde av smittesituasjonen, fordi de færreste nå tester seg når de blir syke.

Ukentlige nøkkeltall publiseres i FHIs ukerapporter.

Periode for grafer: Totalt **Siste 30 d.**

Registrert smittet

93

20. feb. '23 i dag

1 480 136	83	14,9
totalt	nye i går	per 100k siste 14 d.

Trend i reg. smitte

FLAT

SYNKENDE STIGENDE

Flat

Utvikling i trend:

20. feb. '23 I dag

0 kommuner har stigende

Daily data sharing to the public (GitHub) stopped;
automated data processing, aggregation, analysis stopped

Back to manual weekly reports, involving 20+ staff

Risks and Challenges

… or opportunities?

Resources

Funding
Support

Talent

Skills take long time to develop
Attraction

Culture

Collaborative?
Open to new solutions?

Open data and software

Privacy protection related to health data

Correct interpretation of results

Can we trust the **predictions** made by the algorithm?
What does the **uncertainty** mean?
Is there enough **evidence** for an ‘outbreak’?

Robust open source software, maintainence

Training

…

Ways forward

Prevention is better than treatment

Monitoring and surveillance

Infectious diseases in human
Zoonotic diseases (One Health)

Data and information sharing, open source software can make it easier !

Invest in **infrastructure and manufacturing**

Vaccine production and distribution
high protective gears, ⋯

Rapid development of tests and vaccine

Lab security; training and education; ⋯

Stop the spread (at least until vaccine)

2023

?

Resources

Public health surveillance and preparedness

Centers for Disease Control and Prevention (CDC) guide on public health surveillance https://www.cdc.gov/training/publichealth101/surveillance.html

Book on covid and pandemic (“Preventable” by Devi Sridhar)

Coursera course (JHU) on surveillance https://www.coursera.org/learn/epidemiology-surveillance-systems-analysis/

Our world in data https://ourworldindata.org/coronavirus/country/norway

Johns Hopkins COVID data repository https://github.com/CSSEGISandData/COVID-19

FHI data and reports

Weekly report https://www.fhi.no/publ/2020/koronavirus-ukerapporter/

Statistics bank https://statistikk.fhi.no

Sykdomspulsen / CSIDS

Repository of CSIDS (previously Sykdomspulsen) https://github.com/csids

Sykdomspulsen tutorial on autoreport https://github.com/sykdomspulsen-org/resources/blob/main/tutorial_autoreport/autoreport_101.md

Reporting automation https://www.rstudio.com/resources/webinars/rethink-reporting-with-automation/