



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Gesundheit BAG
Direktionsbereich Öffentliche Gesundheit

Building ships while weathering a storm, Arrr...R!

adminR meetup, 18. März 2021

Katrin Schneider / Damir Perisa

Taskforce COVID-19 BAG



Epidemiological surveillance and evaluation

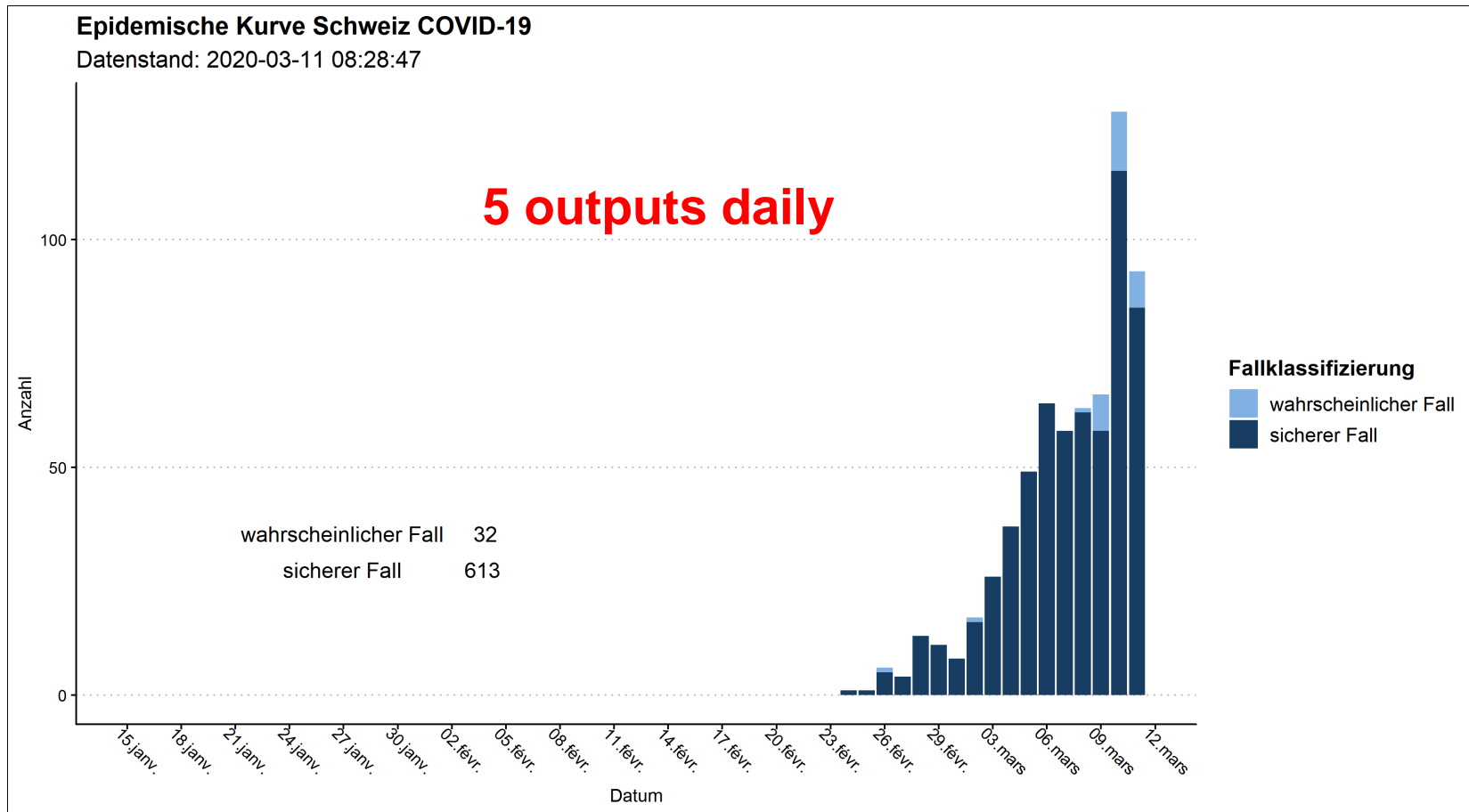
Mission:

To generate, process and assess epidemiological data and information in a timely manner in exchange with internal and external partners.

Produce public health-oriented analyses and reports and communicate them to specific target groups.

AG Lage of the FOPH Taskforce COVID19

A storm is coming ...



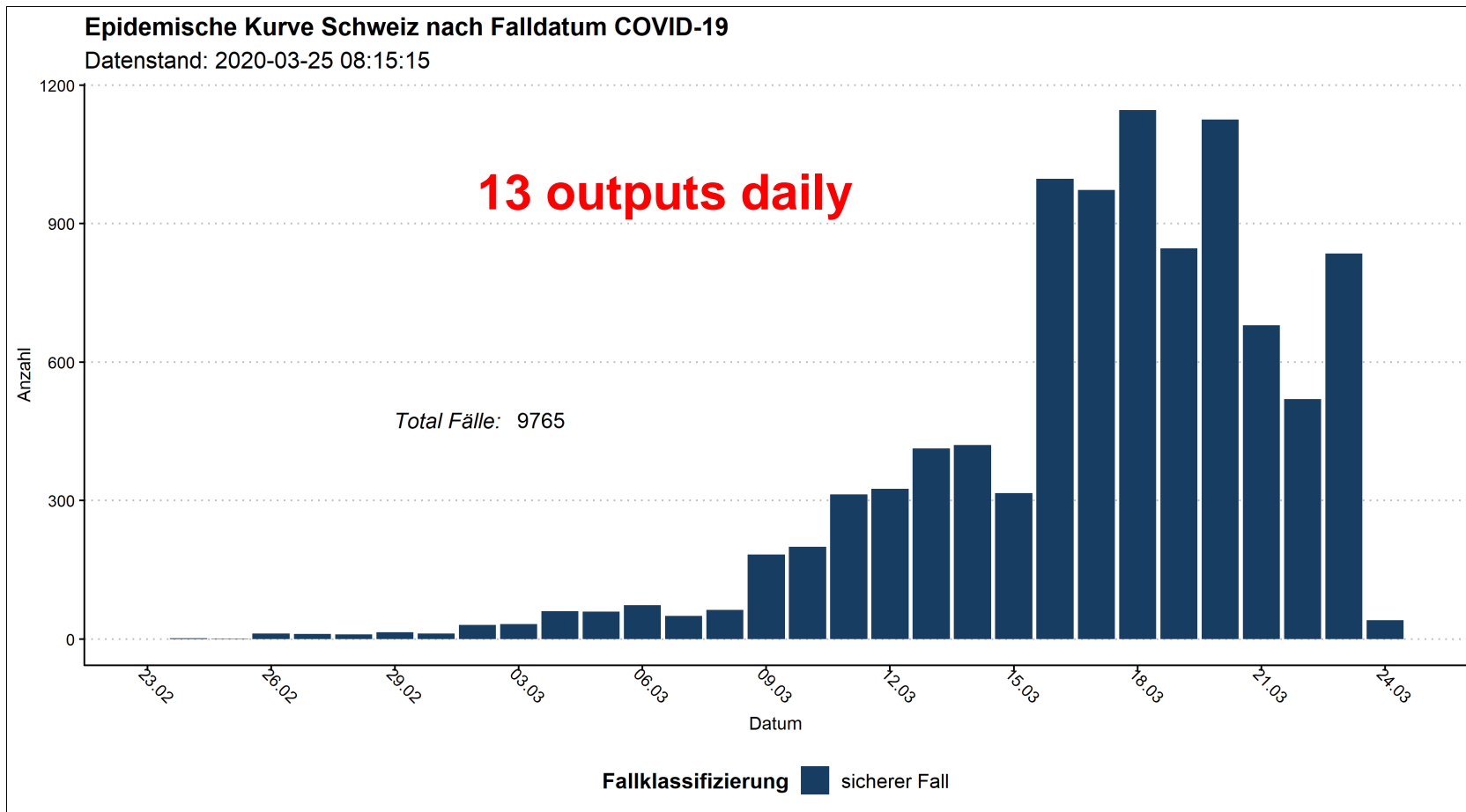
adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



14 days later



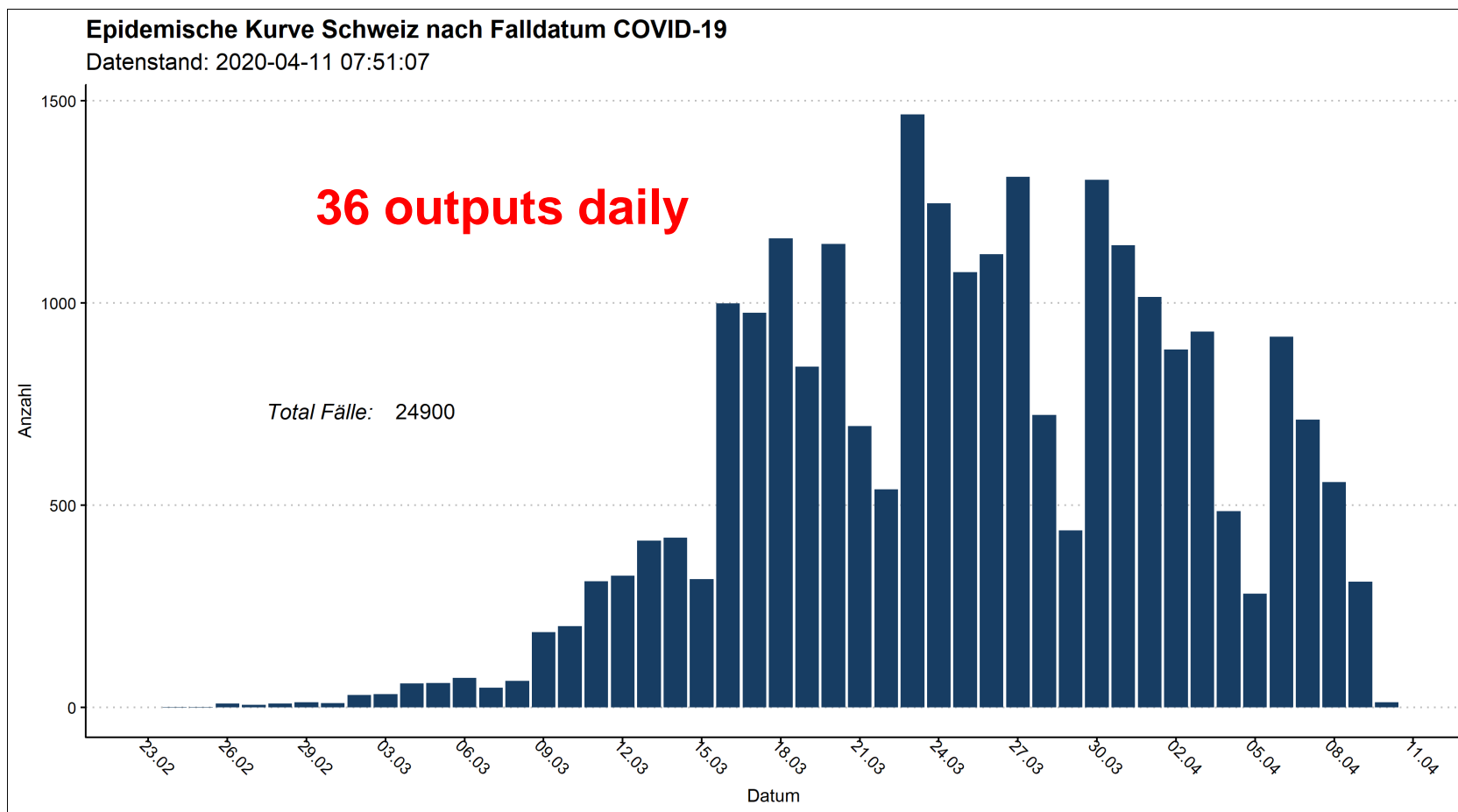
adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



28 days later - case number subsides



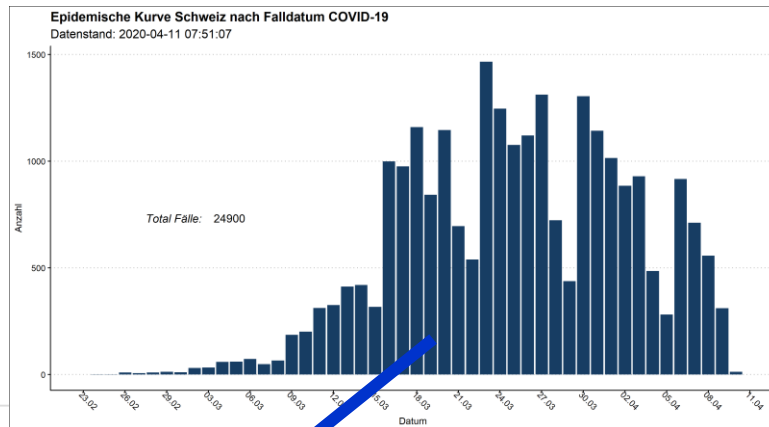
adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



First storm ... was only the beginning!



Fälle pro 100 000 Einwohner/innen

150

100

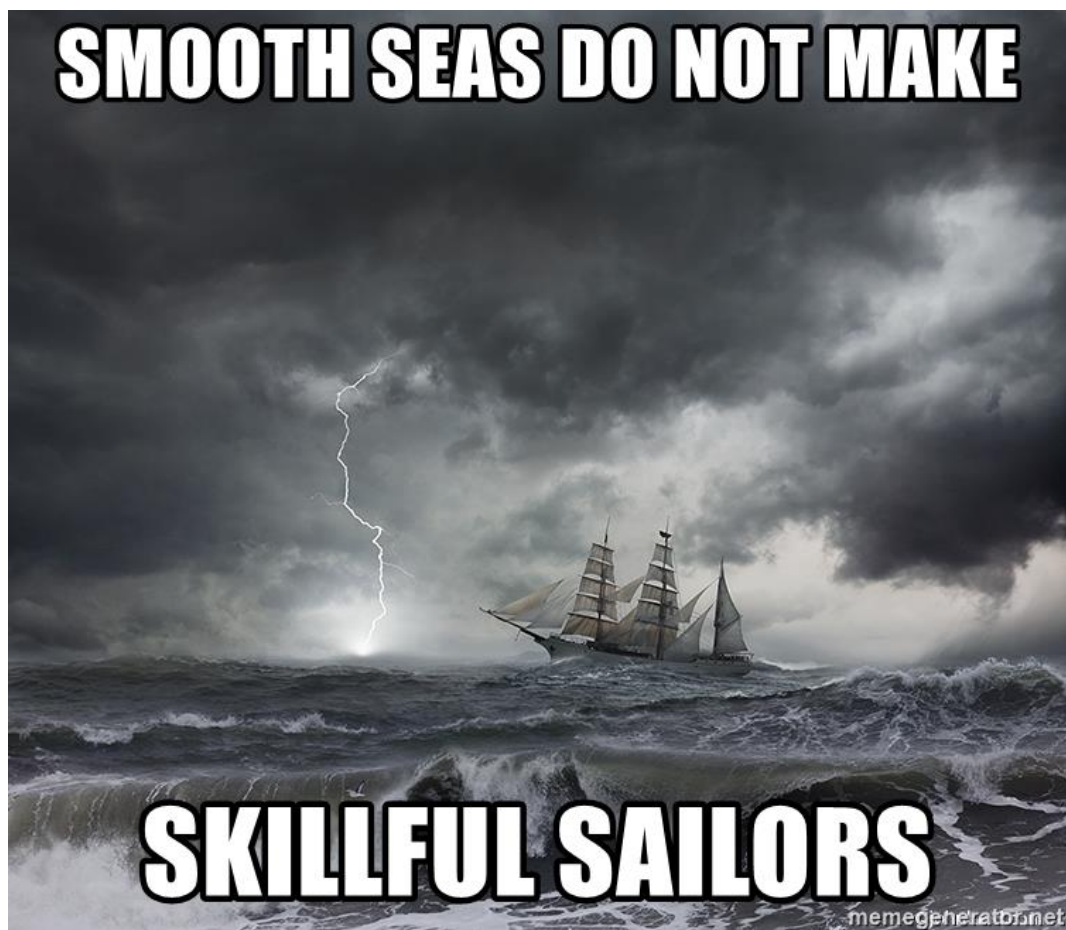
50

0

24.02.2020 01.04 01.05 01.06 01.07 01.08 01.09 01.10 01.11 01.12 01.01 01.02 17.03.2021



Set sails!



adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



Growth over time – cases, data and code

- Januar 2020: 2 R scripts, 54 lines code, runtime **14 seconds**
- Mai 2020: codebase 49 routine scripts (8638 Zeilen) over 30 on demand scripts
- Juli 2020: 63 scripts, 17334 lines code, who knows how many on demand scripts, RAM usage 8 GB
- March 2021: 135 scripts in main stack, 28802 lines code (excluding over 80 foreign and 5 internal r packages), runtime 4425 seconds (**74 min**), RAM usage 23 GB

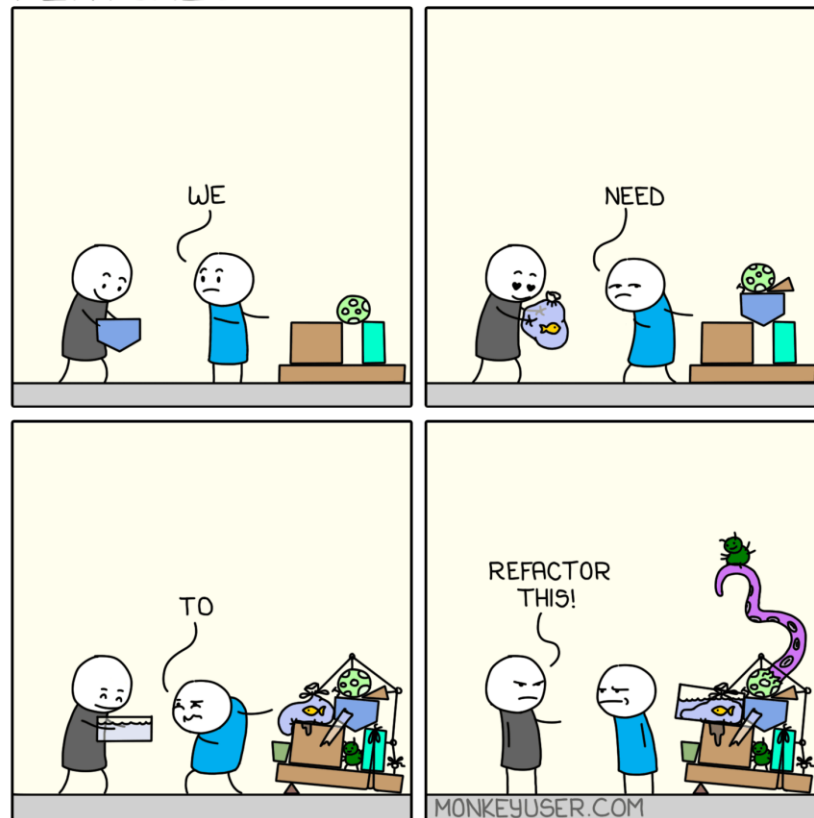




What is taking so long?

- Adding **features** costs
- Adding **data sources** costs too
- **Data** gets bigger daily
 - Code efficiency
 - Database efficiency
 - Transfer times

FEATURES





group_by is expensive – case_when(any()) toooo!

Most cases have none or only
one "kb" entry (= one fall_id)

```
ncov2019_kb <-  
  ncov2019_kb_RAW %>%  
  group_by(fall_id) %>%  
  mutate(  
    var = case_when(any(var_orig == 1) ~ 1,  
                    any(var_orig == 2) ~ 2,  
                    any(var_orig == 3) ~ 3)  
  ) %>%  
  mutate(  
    icu_aufenthalt_kb_merged = if_else(any(icu_aufenthalt == 1), 1L, as.integer(NA)),  
    icu_beatmung_kb_merged = if_else(any(icu_beatmung == 1), 1L, as.integer(NA))  
  ) %>% ...
```



Sustainability in Datascience / Dataoperations

- Manual steps are expensive => Automatisisation
- Benchmarking helps identify pain points
- Team coding
- Reproducibility
 - tools not offering reproducibility are banned (e.g. Excel is forbidden for analysis in production!)
 - All productive code has to be tracked in common git repos (since autumn 2020)
 - Pull requests need at least review of two persons to be accepted

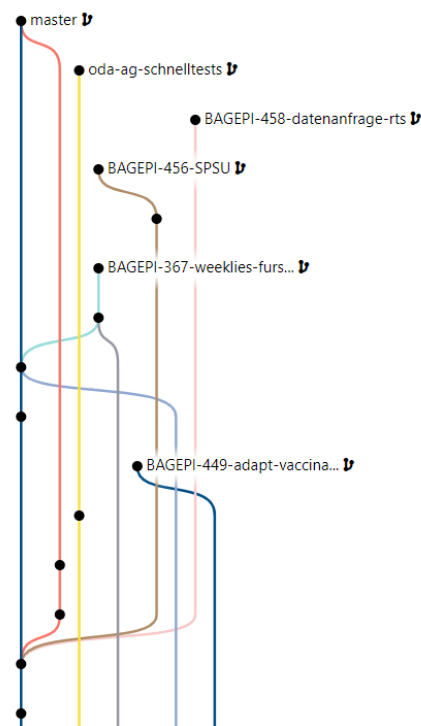


Version tracked code – a life-saver

BAG_EPI / covid19-MadMurdock

All Branches Graph

[All: ☒] ☒ master ☒ bugfix/* ☒ feature/* ☒ h



```
new file: floating_code/QC_impfvariablen.R
modified: floating_code/generate_samples_for_251.R
modified: lib/vaccine_delivery.R
new file: oda/viollier-linkage/Output_for_TStadler/2021-03-15_08-07-18_FOPH_COVID19_dat
a_extract_X_pos_tests_viollier.xlsx
new file: oda/viollier-linkage/Output_for_TStadler/2021-03-15_08-07-18_vis_check_auftra
ggeber_armee.png
new file: oda/viollier-linkage/Viollier_recieved/COVID 19 orders and results von 05.03.
2021 bis 11.03.2021_ETH_BAG.xlsx
modified: oda/viollier-linkage/import_viollier_data.r
new file: oda/viollier-linkage/sequence_report/2021-03-14_bag_sequence_report.csv
new file: tests/data/test_post251_ncov2019_falldetail.rds
new file: tests/data/test_post251_ncov2019_kb.rds
new file: tests/data/test_post251_ncov2019_kb_vars.rds
new file: tests/data/test_pre251_ncov2019_falldetail.rds
new file: tests/data/test_pre251_ncov2019_kb_RAW.rds
deleted: tests/testthat/test-PGM_000_pkg_functions_data_wrangling.r
new file: tests/testthat/test-PGM_008_pkg_functions_data_wrangling.r
modified: tests/testthat/test-PGM_251.R
new file: tests/testthat/test-PGM_964_covid19dboard_VOC.R

U80855744@CM000004 MINGW64 /p/Krisen/2019_nCov/03 AG Lage/07_R/_stash_sources_persoendlich/SCK/cov
id19-madmurdock (BAGEPI-448-im-monitoring-graphs-for-tf|MERGING)
$ git commit -m "merge conflict 372 solved"
[BAGEPI-448-im-monitoring-graphs-for-tf 547859c] merge conflict 372 solved

U80855744@CM000004 MINGW64 /p/Krisen/2019_nCov/03 AG Lage/07_R/_stash_sources_persoendlich/SCK/cov
id19-madmurdock (BAGEPI-448-im-monitoring-graphs-for-tf)
$ git pull origin master
From https://bitbucket.bit.admin.ch/scm/bagepi/covid19-madmurdock
* branch      master      -> FETCH_HEAD
Already up to date.
```

adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021

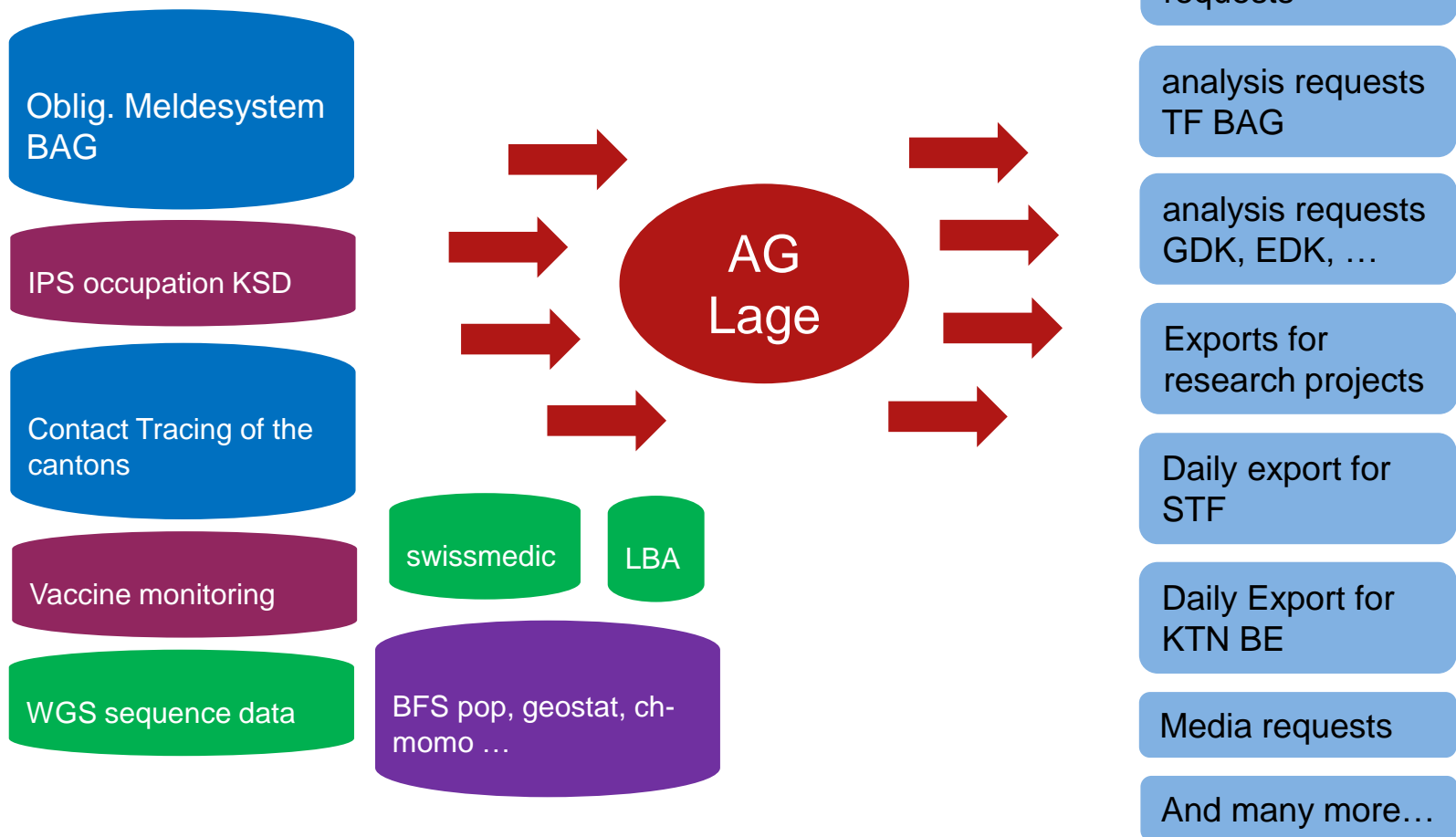


Automatisation is hard

- Identifying areas needing optimisation and refactoring is easy
- Dedicating resources and successfully performing change is hard
- IT environment limits capabilities to automatise
 - Plethora of authentication methods
 - Plethora of encryption methods
 - Rules made for 1990 tools and ways of work
 - Firewall and Proxy issues



Where we are now



adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



Use the right (available) tools





What's next?

Modernisation of toolsets

Harvester DBs

Data Marts

VMs with Docker

continuous integration on all code / Jenkins is our friend!

Sustainability strengthening

crisis mode -> routine mode

Process optimisation

set up collaborations and synergies with cantonal and federal offices

adminR: Building ships while weathering a storm, Arrr...R!

Katrin Schneider / Damir Perisa

18. März 2021



Acknowledgements

Team Datascience / Dataoperations AG Lage :

Ekkehardt Altpeter, Salome Bitter, Jean Delacoste, Reto Bürgin,
Anne Laube, Urs Mayr, Serge Zaugg, Samuel Colin, Jeffrey Keller,
Julien Riou, Nanina Anderegg

BAG Msys :

Marcel Brunner, Marc Filou, Patricia Waldvogel, Andreas Garbin and colleagues

Dashboard Team :

Walter Grolimund BFS

Shiftcode: Patrick Suter, Simon Mummenthaler, Simon Guggisberg,

Simplificator: Florian Zinggeler, Dimitar Petrov, Thomas Plüss

Interactive Things: Tanja Boa, Solange Vogt

BAG IT :

Roger Bussard, Ilir Iseini, Bruno Blatter, Daniel Megert and colleagues

BIT :

Stefan Leisi, Seb Mathyas, Gemmi Nicoletti and colleagues



Dashboard covid19.admin.ch

Demo time!