

Data is the new gold: Information flow during the 2014-2016 Ebola outbreak and the Coronavirus outbreak

Delphine Courvoisier

University of Geneva & Geneva University Hospitals

Pillars of epidemics response

1. Treating sick patients
2. Community engagement
3. Safe burial
4. Surveillance
 1. Information management

Outline

Epidemics, pandemics, and information flow

- A blood sample is not a person
- What's behind the R_0 ?
- Garbage In, Garbage Out (GIGO)
- Infrastructure of data flow, data security,
→ who owns the data
- The example of COVID19 cases and contact tracing in Geneva

A blood sample is not a person

Labs provide the results of each test.

- The quality of the tube's tag is crucial
- An example with Ebola, a week in January 2015
 - CDC (US): 82 cases
 - WHO: 51 cases
- What happened
 - Same source of data
 - Different algorithms

Different algorithm for date of onset

CDC used

- if dead (StatusAsOfCurrentDate), dateOnset=date of death-4
- if dead and no date of death, dateOnset=SampleTest-5
- if dead and no date of death and no SampleTest, dateOnset=date Report-4
- if alive and no dateOnset, dateOnset=DateReport-4
- if alive and no dateOnset and no DateReport, dateOnset=SampleTest-7
- if missing StatusAsOfCurrentDate and no dateOnset, dateOnset=DateReport-4
- if missing StatusAsOfCurrentDate and no dateOnset no DateReport, dateOnset=SampleTest-6
- if DateOnset < 01/05/2015 (May 1st 2015), dateOnset=DateReport-4

WHO used:

- if DateOnset is missing, DateOnset=dateReport-4
- if DateOnset and dateReport are missing, dateOnset=DateDeath-5

We suggest that the procedure for imputing DateOnset should be decided as a policy by the leads of CDC and WHO teams.

Different algorithms for district

District name (DistrictRes or DistrictOnset variables):
these district names are entered using various spelling resulting in more than 70 different values

- CDC used: district of onset, and correct the spelling.
- WHO used: district of residence and if missing, districtRes=districtOnset.

What's behind the R_0 ?

$$\mathcal{R}_0 \propto \underbrace{\left(\frac{\text{infection}}{\text{contact}} \right)}_{\text{Transmissibility}} \cdot \underbrace{\left(\frac{\text{contact}}{\text{time}} \right)}_{\text{Rate of contact}} \cdot \underbrace{\left(\frac{\text{time}}{\text{infection}} \right)}_{\text{Duration of infectiousness}}$$

- **Transmissibility** (ie probability of infection given contact between a susceptible and infected individual)
- Average **rate of contact** between susceptible and infected individuals
- **Duration of infectiousness**

What's behind the R_0 ?

$$\mathcal{R}_0 \propto \left(\frac{\text{infection}}{\text{contact}} \right) \cdot \left(\frac{\text{contact}}{\text{time}} \right) \cdot \left(\frac{\text{time}}{\text{infection}} \right)$$

- **Transmissibility**: may decrease when temperature increases
- **Rate of contact**: underlies decisions for quarantine, confinement, self-isolation
- **Duration of infectiousness**: highlights the importance of finding a drug that would reduce disease duration

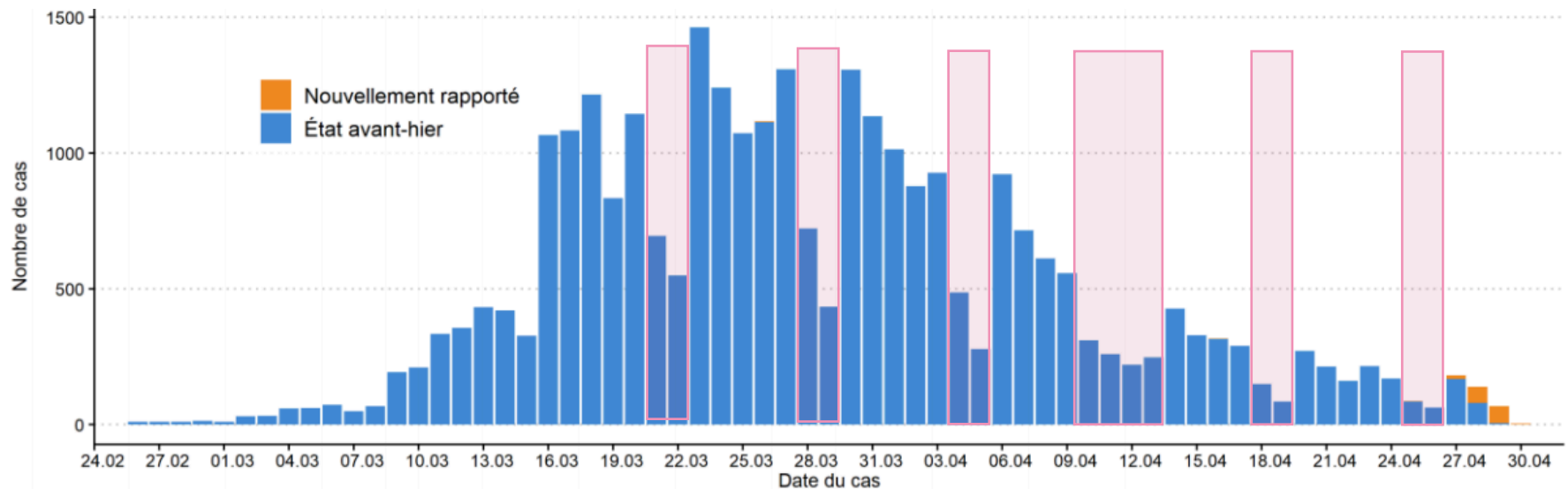
Garbage In, Garbage Out (GIGO)

- Data on patients with the disease or on the number of tests is necessary for:
 - Planning care resources
 - Helping patients
 - Monitoring disease progression and understanding when the disease is contained
- Data quality input is crucial

Garbage In, Garbage Out (GIGO)

Aim	Examples of GI	Examples of GO
Planning care resources	<ul style="list-style-type: none">- SL: missing name info- CH: Fax sent to FOPH may take time to be inputted	<p>Samples were considered duplicates</p> <p>Less precise estimation of R_0</p>

Switzerland epidemiological curve



Garbage In, Garbage Out (GIGO)

Aim	Examples of GI	Examples of GO
Planning care resources	<ul style="list-style-type: none">- SL: missing name info- CH: Fax sent to FOPH may take time to be inputted	<p>Samples were considered duplicates</p> <p>Less precise estimation of R_0</p>
Helping patients	<ul style="list-style-type: none">- GE: Labs do not provide accurate phone number	<ul style="list-style-type: none">- Delay in contacting pts → isolation starts later

Garbage In, Garbage Out (GIGO)

Aim	Examples of GI	Examples of GO
Planning care resources	<ul style="list-style-type: none">- SL: missing name info- CH: Fax sent to FOPH may take time to be inputted	<p>Samples were considered duplicates</p> <p>Less precise estimation of R_0</p>
Helping patients	<ul style="list-style-type: none">- GE: Labs do not provide accurate phone number	<ul style="list-style-type: none">- Delay in contacting pts → isolation starts later
Monitoring disease progression and understanding when the disease is contained	<ul style="list-style-type: none">- Only deaths at hospital are reported- GE: Addresses often wrong or missing	<ul style="list-style-type: none">- Underestimation of deaths and case fatality rate- Incorrect reporting of clusters

Who owns the data?

- Patients own their data! Yes, but:
 - In the case of an epidemics, healthcare professionals need to have access to these data even if the patient has not given consent.
 - Ebola SL: data was hosted in the US because CDC provided a solution for case monitoring first.
 - Law of epidemics, art 12, obligation to declare.
 - To paraphrase Alain Berset: “Security should be as strong as possible, but as lax as necessary.”
 - COVID Geneva: Office of the surgeon general required that the labs provide cellphone numbers to allow rapid monitoring and self-quarantine advice to new cases

Actionnable Registry of the Geneva Outpatient with SARS-Cov2 (ARGOS) - Data flow

Delphine Courvoisier, Denis Mongin, Adriana
Uribe, Jean-Luc Falcone, Sophia Vignard

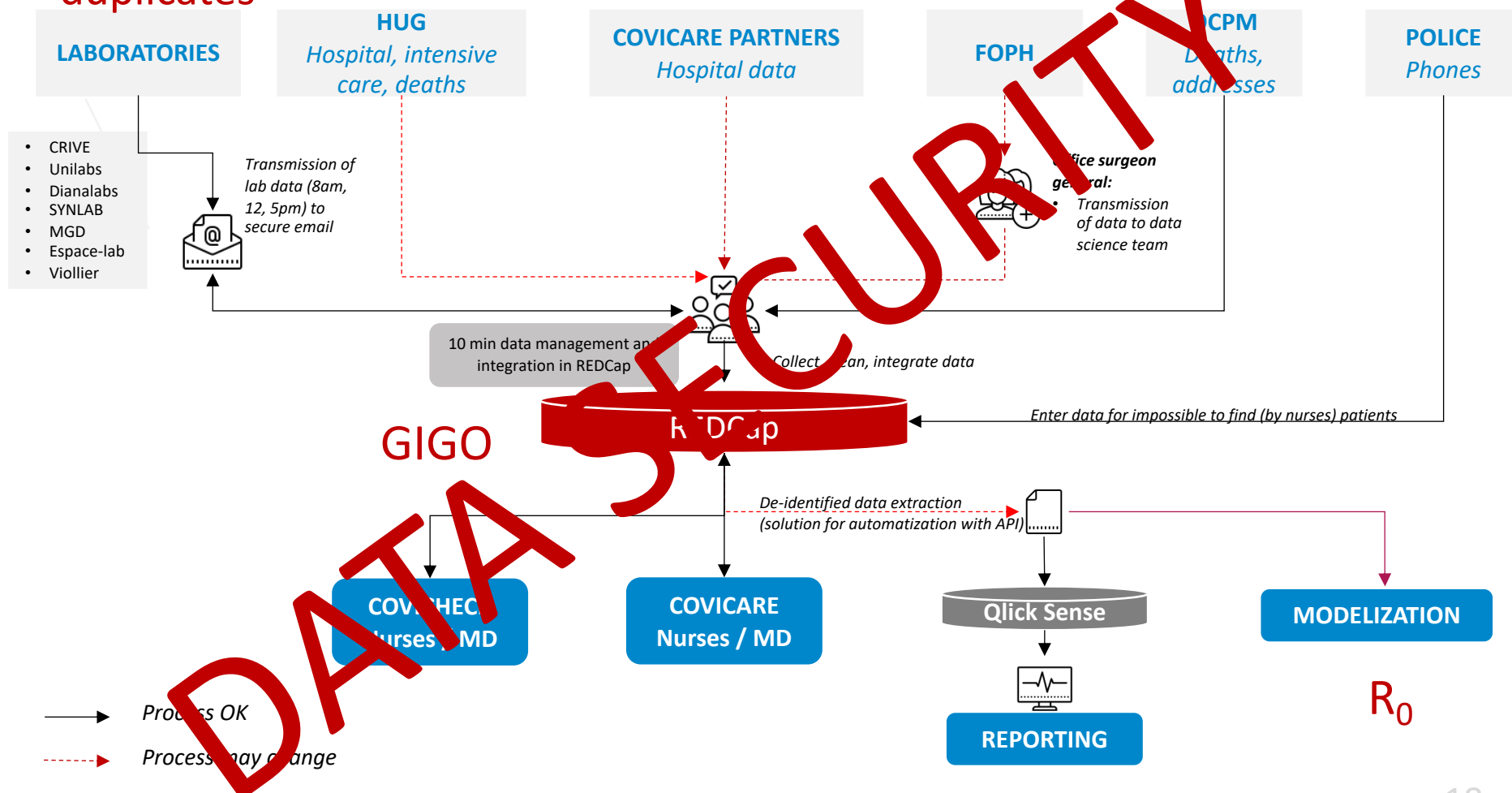
Delphine.Courvoisier@hcuge.ch

ARGOS

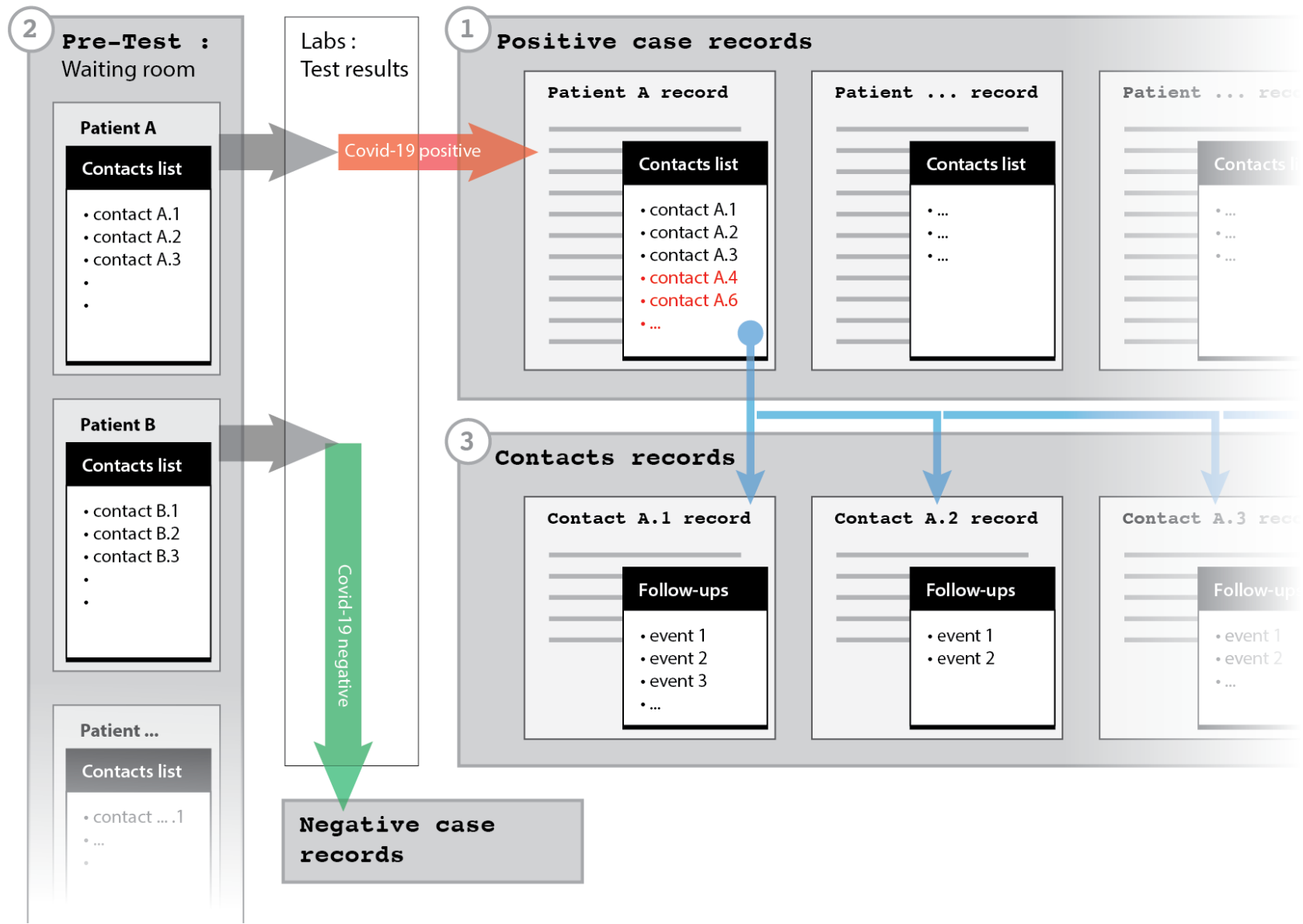
- A blood sample is not a person
→ avoiding duplicates
- What's behind the R_0 ?
- Garbage In, Garbage Out (GIGO)
- Infrastructure of data flow, data security,
→ who owns the data

Data flow: cases

Avoiding
duplicates

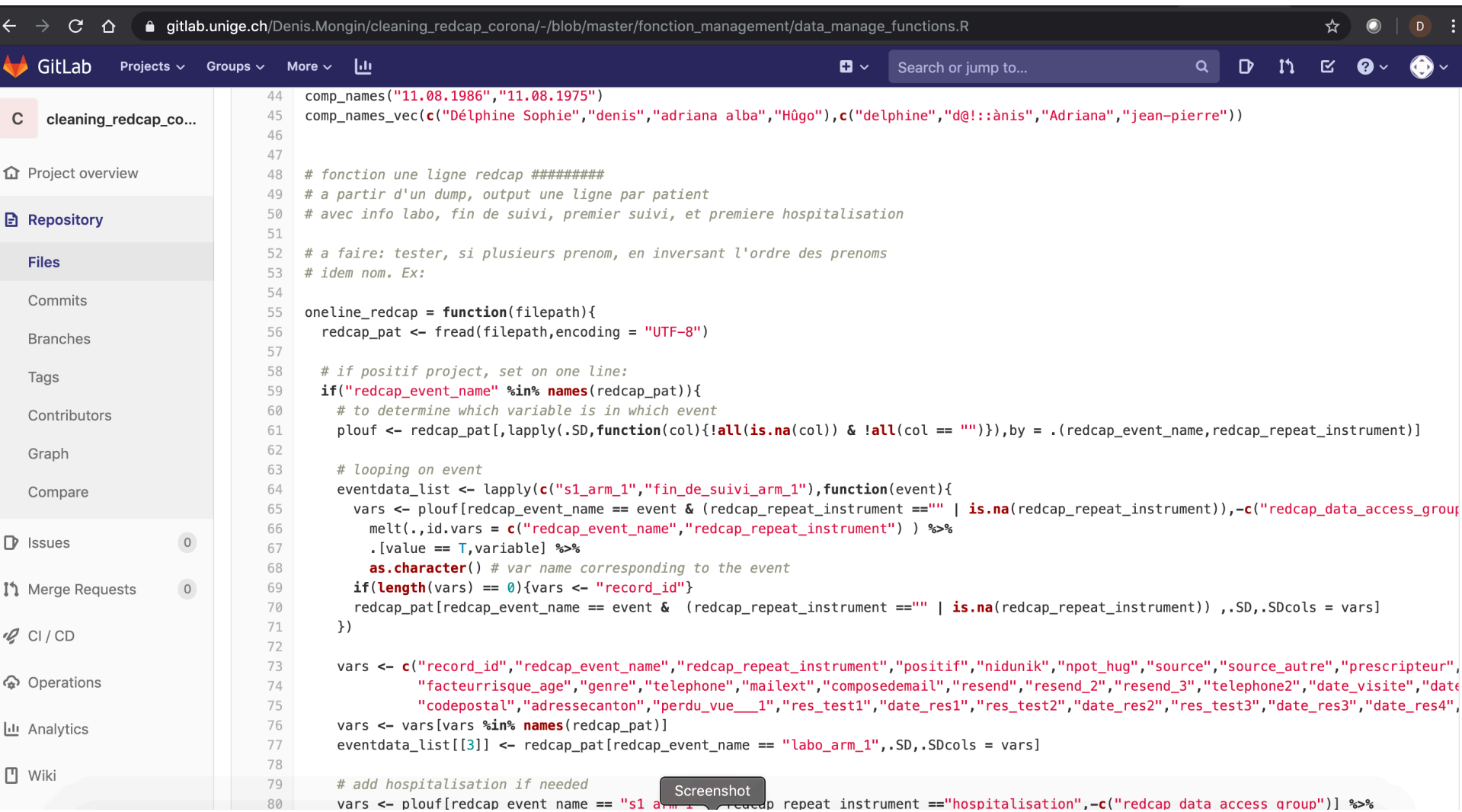


Data flow: contacts



What it really looks like...

~5000 lines of code developed in 6 weeks



The screenshot shows a web browser displaying a GitLab repository page. The browser's address bar shows the URL: `gitlab.unige.ch/Denis.Mongin/cleaning_redcap_corona/-/blob/master/fonction_management/data_manage_functions.R`. The GitLab interface includes a sidebar on the left with navigation links: Project overview, Repository, Files, Commits, Branches, Tags, Contributors, Graph, Compare, Issues (0), Merge Requests (0), CI / CD, Operations, Analytics, and Wiki. The main content area displays the R code file `fonction_management/data_manage_functions.R`. The code is written in R and includes comments in French. It defines a function `oneline_redcap` that takes a `filepath` as input and returns a data frame. The function uses `fread` to read the data, processes it based on the `redcap_event_name`, and returns a data frame with columns `record_id`, `redcap_event_name`, `redcap_repeat_instrument`, `positif`, `nidunik`, `npot_hug`, `source`, `source_autre`, `prescripteur`, `facteur_risque_age`, `genre`, `telephone`, `mailext`, `composedemail`, `resend`, `resend_2`, `resend_3`, `telephone2`, `date_visite`, `date_visite_2`, `codepostal`, `adressesecanton`, `perdu_vue_1`, `res_test1`, `date_res1`, `res_test2`, `date_res2`, `res_test3`, `date_res3`, `date_res4`, and `date_res5`. The code also includes a comment about adding hospitalisation if needed.

```
44 comp_names("11.08.1986","11.08.1975")
45 comp_names_vec(c("D  lphine Sophie","denis","adriana alba","H  go"),c("delphine","d  !:  nis","Adriana","jean-pierre"))
46
47
48 # fonction une ligne redcap #####
49 # a partir d'un dump, output une ligne par patient
50 # avec info labo, fin de suivi, premier suivi, et premiere hospitalisation
51
52 # a faire: tester, si plusieurs prenom, en inversant l'ordre des prenom
53 # idem nom. Ex:
54
55 oneline_redcap = function(filepath){
56   redcap_pat <- fread(filepath,encoding = "UTF-8")
57
58   # if positif project, set on one line:
59   if("redcap_event_name" %in% names(redcap_pat)){
60     # to determine which variable is in which event
61     plouf <- redcap_pat[,lapply(.SD,function(col){all(is.na(col)) & !all(col == "")}),by = .(redcap_event_name,redcap_repeat_instrument)]
62
63     # looping on event
64     eventdata_list <- lapply(c("s1_arm_1","fin_de_suivi_arm_1"),function(event){
65       vars <- plouf[redcap_event_name == event & (redcap_repeat_instrument == "" | is.na(redcap_repeat_instrument)),-c("redcap_data_access_group")]
66       melt(.,id.vars = c("redcap_event_name","redcap_repeat_instrument") ) %>%
67         .[value == T,variable] %>%
68         as.character() # var name corresponding to the event
69       if(length(vars) == 0){vars <- "record_id"}
70       redcap_pat[redcap_event_name == event & (redcap_repeat_instrument == "" | is.na(redcap_repeat_instrument)) ,.SD,.SDcols = vars]
71     })
72
73     vars <- c("record_id","redcap_event_name","redcap_repeat_instrument","positif","nidunik","npot_hug","source","source_autre","prescripteur",
74             "facteur_risque_age","genre","telephone","mailext","composedemail","resend","resend_2","resend_3","telephone2","date_visite","date_visite_2",
75             "codepostal","adressesecanton","perdu_vue_1","res_test1","date_res1","res_test2","date_res2","res_test3","date_res3","date_res4","date_res5")
76     vars <- vars[vars %in% names(redcap_pat)]
77     eventdata_list[[3]] <- redcap_pat[redcap_event_name == "labo_arm_1",.SD,.SDcols = vars]
78
79     # add hospitalisation if needed
80     vars <- plouf[redcap_event_name == "s1_arm_1" & redcap_repeat_instrument == "hospitalisation",-c("redcap_data_access_group")] %>%
```

A multidisciplinary team effort

Thank you!



Nurses
Doctors
IT team
Data scientists
Biologists

Doctors / Nurses
Digital health experts
Epidemiologists
IT team / Business intelligence
Project managers
Business analysts
Police officers

IT developers
Data scientists