A note on standards and applications for automatic contact tracing.

Prof. Axel Legay (PhD), Ecole Polytechnique de Louvain, institut des technologies de l'information et de la communication de l'électronique et des mathématiques appliquées (ICTEAM), Département d'informatique, Secteur des Sciences et Technologies, UCLouvain axel.legay@uclouvain.be

Prof. ord. Jean-Luc Gala (Md, PhD), Centre de Technologies Moléculaires Appliquées (CTMA), Institut de Recherche expérimentale et Clinique (IREC), Secteur Sciences de la Santé, UCLouvain jean-luc.gala@uclouvain.be

A. Summary:

Over the last few days, there have been several discussions regarding the deployment and use of web applications for "automatic contact tracing" among the Belgian population. This short document is a snapshot document presenting, explaining and summarizing some very relevant observations to be considered when discussing this technology. It is indeed essential for decision-makers to understand what is hidden behind some applications and what is at stake in terms of efficacy of the application and the risk related to its use in terms of personal data protection. This should enable them to choose for the most efficient and less intrusive options in terms of successful automatic tracing protection of data privacy, respectively.

It is of note that many citizens have a biased view of what such an application should be. Actually, citizens are scared that national authorities could observe/control them because the application is based on the private content of their own cellular phone. They should however keep in mind that all the information that they put daily on Google and Apple when using various applications (e.g. online shopping, phone localisation ...) is as vulnerable as the "automatic contact tracing app" discussed hereafter in this synthetic document.

As expert scientist in this type of technology we are therefore puzzled by the reluctance of the citizens to use a "contact tracing application", whereas they so easily accept to use daily other e apps that are notably far more intrusive.

The objective of this document is not to plead for a specific application but to inform the citizens so that they can understand the issues related to an "automatic contact tracing", and understand whether using some proposed solutions is safe or rather unsafe for their "own personal data privacy".

To reach this goal, a set of solutions will be equally compared so that citizens are offered the chance to be correctly informed, hence to give the best informed consent. Consequently, this note considers privacy protection but goes beyond it.

Table des matières

B. Frequently asked questions:

What is the objective of a contact tracing?

Such application aims to "rapidly" and "automatically identify all recent contacts" of an individual becoming suddenly "positive" due to the contamination by the SARS-CoV-2.

What do we mean by "recent contacts" of a positive patient?

We know about the "viral shedding", which means the identification of the infection (viral genome) in the respiratory tract specimens of a positive patient (hence its potential for viral transmission to close contacts), may start to 1–2 days before the onset of symptoms!

Therefore, choosing for optimal safety and health preservation within our society would mean that we monitor all the contacts over the last two or three days preceding the onset of symptoms in a confirmed case. Contacts of this confirmed case are then automatically warned that they have possibly had an infectious contact two to three ago time with a positive case (identity of the latter is never disclosed!), and that they are therefore requested to carefully monitor their health status over the next two weeks (what we call the incubation time: on average 4 to 5 days for most individual, but with extreme duration going from 1 to 14 days). The contact of a confirmed positive case is then requested to consult a medical doctor and to be tested for SARS-CoV-2 infection (what we call the presence of a viral load in the nasopharyngeal track, as tested by nasal swabbing followed by DNA-based testing), in case symptoms are compatible with a viral infection.

Bluetooth protocol

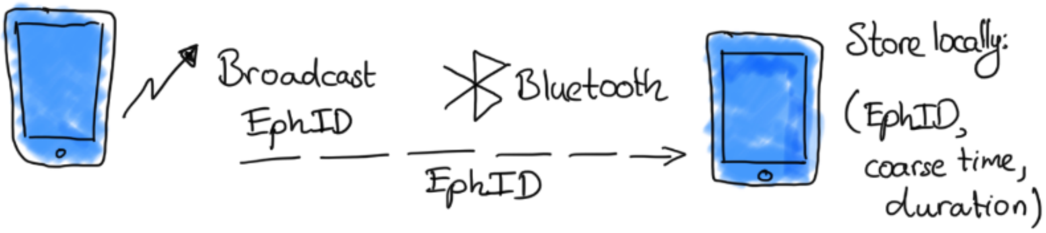
Most of contact tracing solutions use the *Bluetooth protocol*. This protocol aims at sending *Ephemeral Identification Numbers* (*EphID*) to establish a relationship between two or more connected objects, like smartphones. The *EphID* is meant to be anonymized, and be used without disclosing any geographical nor personal, private information of the smartphone's owner.

Note that almost everybody already uses the Bluetooth protocol (e. g. cars and headphones) and hence already and may surely propagate many extremely private and personal information all around the world, without even being aware of it and –without being ever informed about the content, and about related risks!

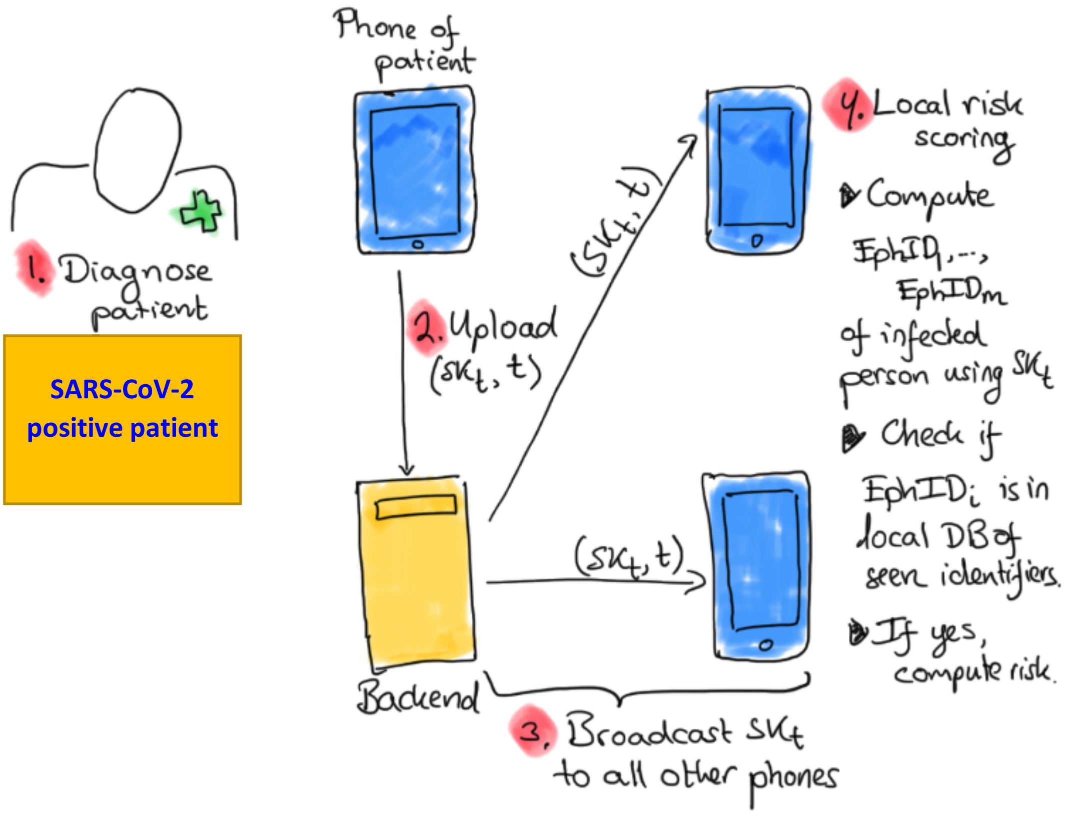
Decentralized contact tracing

Among existing approaches for an "*automatic contact tracing*", the DP-3T (https://github.com/DP-3T/) proposed standard has raised many questions that are summarized hereafter. In a nutshell, the standard works as follows.

- **Step1, local recognition between phones:** Each phone exchange random number (EphID) with other phones. Those numbers are recorded and represent your ID (not your phone number nor your position and even less your personal identity) for other phones during a specific amount of time. It means that you keep the same EphID for a specific amount of time and then a new ID is generated (for example every 15 minutes). We also record the pairs (EphID - time) during a specific period. This period depends on the opinion on experts regarding the risk and occurrence of viral shedding in a symptomatic positive patient. We know that the viral shedding can occur **one to (maximum) two days** before the onset of symptoms in a SARS-CoV-2 positive patient)).



- **Step 2, communication phase:** In case a user gets sick (1, diagnose patient), he/she sends his/her ID to a central server. The server uploads the random number and the time when it was generated (point two in the figure below). The whole process takes place in an entirely anonymous way with strong cryptographic protocols that are detailed in (https://github.com/DP-3T/). The central server sends the number to all phones that are using the application (points 3 and 4 on the figure below). Each phone checks whether it has ever met the number in the last two days (if a two-day interval is selected as the best time interval for automatic tracing). In case of a match, a local risk algorithm is run in order to see whether there is a risk that this person was contaminated by the infected person.


Phone of patient
1. Diagnose patient
SARS-CoV-2 positive patient
2. Upload (SK_t, t)
(SK_t, t)
(SK_t, t)
Backend
3. Broadcast SK_t to all other phones
4. Local risk scoring
Compute EphID_1, ..., EphID_m of infected person using SK_t
Check if EphID_i is in local DB of seen identifiers.
If yes, compute risk.

In the remaining part of this note, and considering that DP-3T is an open source application, the focus is set on DP-3T efforts to prevent a lack of transparency as well as a possible underestimation of current known vulnerabilities. What we observed is that the authors carry out openly and in a fully transparent way a fair comparison between their app and other existing solutions.

As external observer, fully independent from the DP-3T consortium, we do really appreciate the commitment of these DP3-T top scientists to full scientific transparency. We are convinced that what they propose deserves discussions and debates, and not prior assessment critics, bashing and rejection in principle.

Credits: The source of all the pictures from the present document is the open source DP-3T website.

Centralized contact tracing

There are other open source solutions like the PEPPT-PT protocol. This solution is centralized, hence gives more power to a central authority. The main difference between EphID numbers and a decentralized solution is that, in the centralized option, EphID numbers are recorded at the level of a central server (often supervised and controlled by state authorities), and this server is precisely the one informing all the contacts who have met a confirmed positive case the one or two preceding days. The central server is also likely to generate the ID, which increases the risk of major detrimental should a potential attacker (i.e. a person who wants to hack the solution) succeed to take control of the server.

- ## What is the opinion of the national commission for private life protection?

It is worth mentioning that albeit many claims that contact tracking would break private life (see e.g. , the CNIL (*Commission Nationale de l'Informatique et des Libertés*) conditionally agreed to the development of StopCovid, the French application for contact tracing. See https://www.cnil.fr/fr/publication-de-lavis-de-la-cnil-sur-le-projet-dapplication-mobile-stopcovid .

StopCovid is centralized as described above. The go is conditioned by the fact that

1. Uploading the application is not mandatory; it remains a free initiative and citizen's decision to download it or not and based on a transparent and understandable information on how data will be used and stored),

2. The application is developed in the context of a global response to the COVID-19 pandemic and use solely for this purpose.

- ## Implementation challenges

The above protocols are standards that will be implemented in concrete applications. Accordingly, related technical challenges are:

  - To define cryptographic protocols and random generators of EphIDs to guarantee privacy.
  - To make sure that the Bluetooth protocol can be accessed as background.

Both DP-3T and PEPPT-PT standards provide solutions to the first point and are very transparent on their respective genuine and practical limitations.

Apple and Google are providing solutions to the second technical challenge, as owner of the operating system. This means they have the control on the Bluetooth setting. Whereas the appropriate setting has to be activated by them, both Apple and Google have committed to provide a solution to this challenge by mid-May. It is noteworthy that anyone can already implement/test applications with the actual Bluetooth configuration. However, as it will run in front end, this will consume much more energy.

- **Is the Bluetooth technology ready for tracing applications? Is it true that it will generate many false-positive results (a hypothesis that would undoubtedly be alarming and anxiety-producing)?**

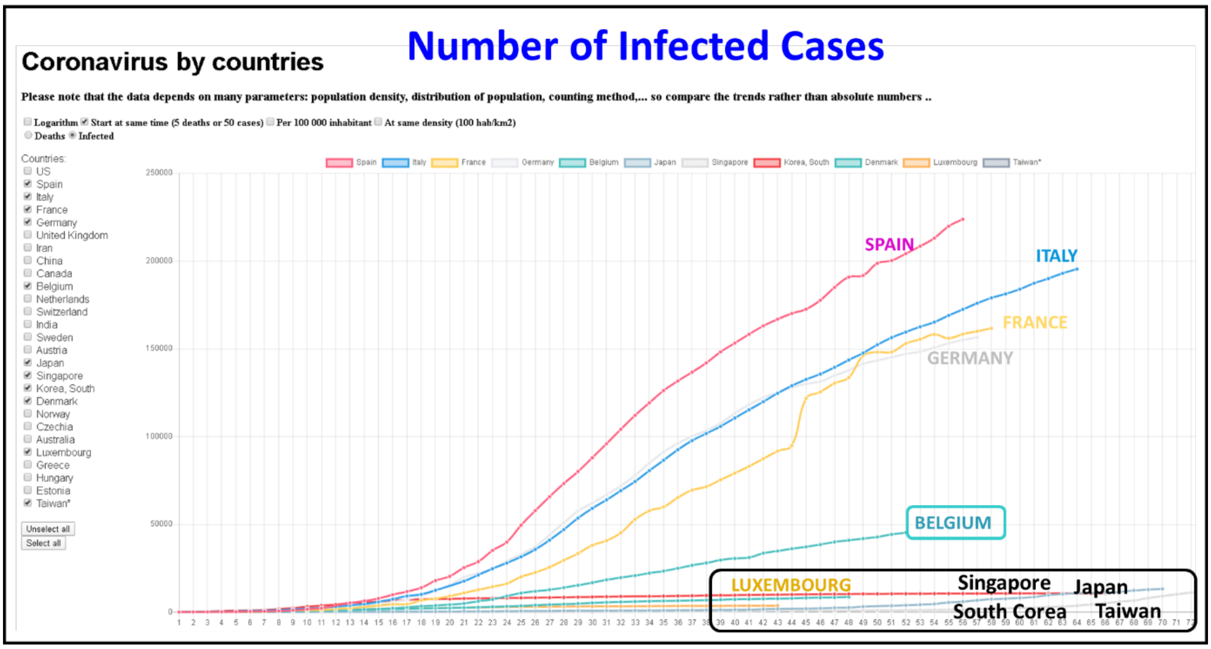
Indeed, the Bluetooth area (i.e. the area where the signal is active) varies from few to hundred meters. In addition, Bluetooth is not limited to open areas but can cross walls. In that respect, we need to consider the following practical situations:

- If you're sitting in a bus at a distance of two meters from an infected individual, the algorithm will indeed conclude, when later activated at the time of positive case confirmation, that you **were maybe in contact** with this positive case. The risk analysis shall only provide a warning to help you and keep you alert about this potential risk for your own health, but you will still need to call a doctor to get more advices and you will need to carefully monitor during the whole incubation period (from 1 to 14 days with an average of 4 to 5 days from most individuals).

- The same applies in case of suspicious contact in a building or any other closed area;

- **Do those applications have a chance of success in Belgium when used for an "automatic contact tracing" during the current COVID-19 pandemic, and more particularly, during the deconfinement phase?**

It is true that only 16% of Singapore population used this application in the deconfinement phase. It is, however, of note that it represents nearly 1 million citizens, and worth to consider that Singapore recorded a lower fatality rate with a slower viral progression than in Belgium.



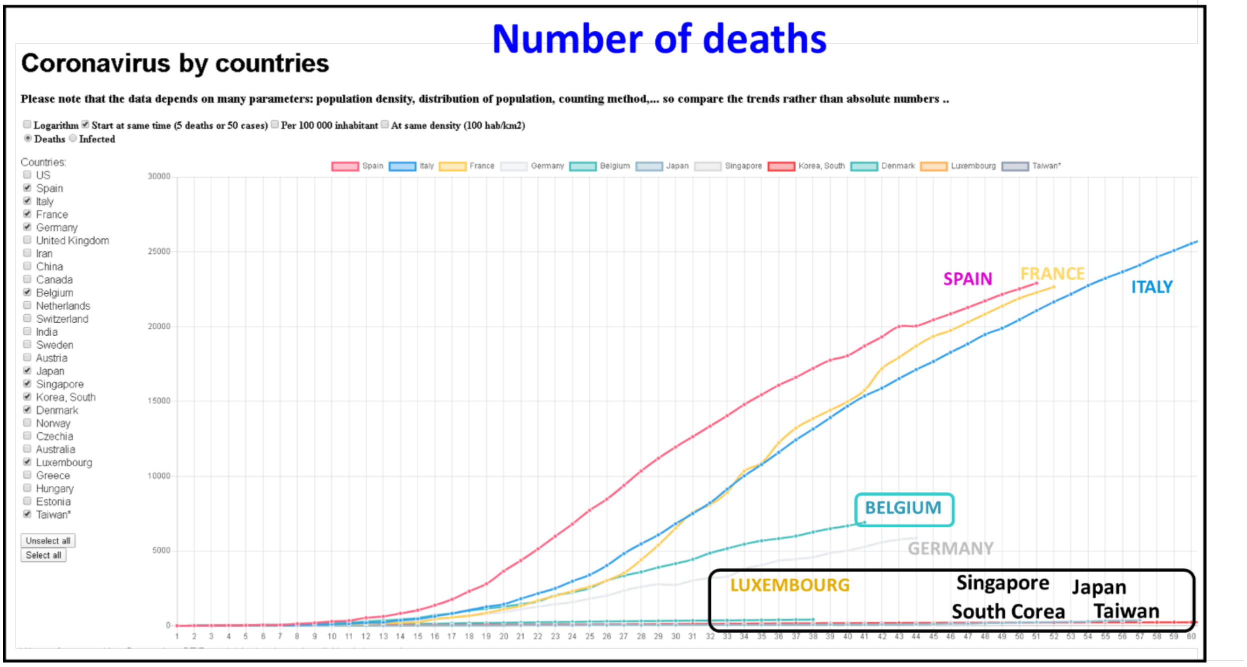
Number of Infected Cases

Coronavirus by countries

Please note that the data depends on many parameters: population density, distribution of population, counting method,... so compare the trends rather than absolute numbers ..

☐ Logarithm ☑ Start at same time (5 deaths or 50 cases) ☐ Per 100 000 inhabitant ☐ At same density (100 hab/km2)
○ Deaths ● Infected

Countries:
☐ US
☑ Spain
☑ Italy
☑ France
☑ Germany
☐ United Kingdom
☐ Iran
☐ China
☐ Canada
☑ Belgium
☐ Netherlands
☐ Switzerland
☐ India
☐ Sweden
☐ Austria
☑ Japan
☑ Singapore
☑ Korea, South
☑ Denmark
☐ Norway
☐ Czechia
☐ Australia
☑ Luxembourg
☐ Greece
☐ Hungary
☐ Estonia
☑ Taiwan*



Coronavirus by countries

Please note that the data depends on many parameters: population density, distribution of population, counting method,... so compare the trends rather than absolute numbers ..

□ Logarithm ☑ Start at same time (5 deaths or 50 cases) □ Per 100 000 inhabitant □ At same density (100 hab/km2)
● Deaths ○ Infected

Countries:
□ US
☑ Spain
☑ Italy
☑ France
☑ Germany
□ United Kingdom
□ Iran
□ China
□ Canada
☑ Belgium
□ Netherlands
□ Switzerland
□ India
□ Sweden
□ Austria
☑ Japan
☑ Singapore
☑ Korea, South
☑ Denmark
□ Norway
□ Czechia
□ Australia
☑ Luxembourg
□ Greece
□ Hungary
□ Estonia
☑ Taiwan*

Source: http://phico.io/coronavirus

- ## Do they provide informative information in countries where they were previously used in the same context?

Experts said that 60 to 70% of the population must be immunized in order to break the chain of viral transmission (what we called the “collective immunity”). As an extension, it was said that 60% of the population should use contact tracing for it to be efficient and that the Singapore figure of 16% clearly shows a global inefficiency. This claim only makes sense if considering that the application is the only solution. This is actually not what DP-3T and PEPP-PT (and hence the CNIL) do consider: the application shall be a helper in an integrated global response, which involves a series of complementary counter-measures, but will never be the only solution, and should never be viewed in this way.

- ## What is the difference between European and Asian views on contact tracing?

In Asian countries, this application is mainly used to inform people that have been in contact with infected people. Then, state authorities take over and use another application to control people at home (e.g. by camera and picture, phone calls, and/or geopositioning).

Likewise, the European approach supports the application as a helper to notify people that have been in contact with an infected person. However, we do not promote the usage of this application to control the citizen afterwards, nor to display the names and other personal private data of infected citizens. We, in Belgium, are not control freaks.

The application that we wish to develop is "open source", most of the time decentralized, and totally transparent regarding the genuine technology/human limits. Whether or not to use this application is the sole responsibility of the citizen and his own informed consent. In addition, we would like to point out that this application can only work if it is combined with a transparent testing policy, associated with other complementary measures such as "barrier gestures" (e.g. wearing a mask and keeping social distancing).

• Which European countries do and do not now consider the use of contact tracing?

Here is a partial list (it is of note that this type of national decision is part of a dynamic crisis response plan that may evolves over time:

- **Now** considered for implementation: France, Germany, Austria, Italy, UK, Denmark, and Spain.
- Not considered for implementation: Belgium.

• Who should develop those applications?

Who should develop is not really the key issue. What is crucial is to make sure that the application is interoperable with other applications, i.e., that it relies on a common communication and decentralized communication principle. Some experts wish that the EU coordinates all the national initiatives. Practically speaking, this means that the EU could define standards of application, which are applicable for all applications. This would without any doubt substantially ease "inter-app communication" within the EU.

• Is our personal life and private data adequately protected?

With respect to data protection and safety, we need to keep in mind that any standard discussed above can be threatened and attacked at any time by hackers. Most computer science technologies, including simple websites, are confronted to the same risk: they can all be toned down if hackers decide to launch multiple connections, totally saturating and finally blocking the application.

One of the main questions that has not yet been discussed is the identity of the hacker(s). Centralized solutions make a state hacker more dangerous as he can focus on one system to block, but at the same time, this solution provides a more flexible use for applications. On the contrary, decentralized solutions decrease the hacker capacity to block a whole system but this solution limits the providers in their collect of information. In any case, other users can one day turn into hackers with various type of nuisance related to the system and their own skills and computing resources.

What is really interesting with a contact tracing application is that solution developers stay transparent about probability of a hacking as well as the attack probability related to the hacker resources at private (individual or collective) and state level. Comparisons between the probability of DP-3T and PEPP-PT potential attacks also exist. Consequently, the user is informed about the level of risk and its probability.

From the conducted analysis, one can claim that solutions under discussions are probably more secured than any communication applications (including social networks) used by citizen who are already often releasing unwittingly they private data (localization of phone, roundtrips from google published each month, etc).

• Bashing argumentation

The following argument is often used as last resort by detractors of the technology: “even if the application is safe, it could be combined with other more intrusive applications”. This has been observed in Asian countries or in Poland, but not in Western Europe. Here, we rather view a combination of the application with a human support to guide those who receive a notification that they may have been infected. To the best of our knowledge, there is no numeric control freak extension planned.

• What is the difference with the database recommended by the national council?

It is too early to answer to this question as there are no technical details yet on how the data base will be used, anonymized (or pseudonymized). There is no guarantee that a cross over with other databases will be prevented or avoided. We are moving to a centralized solution with many information from different areas. This is a very complex security problem. The role of human inspectors is also unclear. It is known from the past that merging several databases remains a very hard problem. It is important to point out that the government should also talk to cyber- and app-technicians and not only to lawyers and jurists. The latter’s are useful to protect the rights of the citizens, not to develop this type of app, nor to discuss the risk of use.

How this solution will collect names of people that meet in a public area (like a shop, a park, a museum) is also hard to apprehend. It may be that the proposed governmental solution (not yet implemented at the time of the writing of the current document) will also have to use some numeric methods, e.g., to collect information about people who meet in a public area. Such people remain unknown from the person who will later develop the COVID-19 symptoms, so the only solution may be to record all individuals entering the public area at a specific moment of time. If done in each public area, this approach would actually lead to a similar process than the Bluetooth one, except that names would not be provided by Bluetooth but by other means such as the RFID code. This is actually the solution used in China.

Over the last days, medical experts have expressed their opinion on the use of this database. Some of them seem to compare it with the approach used for tuberculosis. Those doctors are obviously misled because they try to compare the current crisis with their medical background; so they try to compare situations that cannot be because they are simply very different:

- There is no database for tuberculosis in Belgium
- Even if such data base existed, it should be stressed that much more information can be extracted from 50 000 people (the number of current infected COVID-19 cases at the time of writing is 46.700) than from ~1000 patients infected by tuberculosis.
- The usage of the information and disease management is not the same for COVID-19 than for Tuberculosis.

C. Other related questions:

Hereafter, we summarize some other questions received from citizens over the last days.

Q1: Is this document an advocacy in favour of contact tracing?

A1: No. The sole objective of this document is to give a clear view on how those applications work, and to compare them with other existing technologies.

Q2: Why do you want to promote such a document?

A2: To make sure that the population is informed in an open source fashion. To make sure that the population can ask its own questions without being influenced by lobbies.

Q3: Are you in favour of forcing the usage of tracing applications?

A3: No, some people do not even use a phone. A more integrated global strategy (e.g. gesture barriers, tests, risk analysis) implementing a series of complementary SARS-CoV-2 counter-measures and embedding this application used on voluntary basis. It is also important to pinpoint that this global integrated strategy is usefully completed by the collection of data voluntarily generated by the citizens when they answer questionnaires that enable us to monitor the evolution of the pandemic at a sub-local, regional and national scale. Such non-intrusive websites are already available in Belgium.

Q4: Is there a risk that we will eventually be forced to use such application?

A4: As for now, we have given up even the idea of proposing this type of application. We note that such decision leads unavoidably to a restriction of human rights, a price to pay for those who are willing to use this application. We are thus far from the Korean situation where the application is imposed and appears to be truly intrusive (for instance, with pictures at home as a proof of presence and respected containment legal disposition).

Q5: How would you foresee the use of such an application?

A5: The application could only inform you regarding the possibility that you've been in contact with an infected person. It is not meant to control you, to access your personal data (like your ID or your private and work address), nor to take pictures of you. The technology is not perfect, and, as said above, a digital contact does not mean a physical contact. This application is only a helper that shall be combined with test campaigns, automatic call centre, and risk analysis algorithms to help warn the citizen regarding the risk of developing the COVID-19 when he has potentially met a positive case: nothing less, nothing more. People without phones as well as those reluctant to the use of new technologies must also be protected, and so it will be. A combination of various parameters such as mask, technologies, diagnostic tests, treatment and health education on the risk of contagious contacts, are the key to success.

Q6: What is the difference with a GPS-based solution?

A6: There is a difference between tracking and contact tracing solution. Tracking technology uses GPS signals and geographical position. Most applications that were associated with big scandals over the past few years were using tracking information. We should bear in mind that an inspector interviewing the contact of a positive individual will very easily know all geographical information related to him. On the contrary, contact tracing gathers anonymised local information but does not need to access your geographical data.