

The Inria logo is written in a white, elegant cursive script on a solid red rectangular background.

Les Rencontres de Santé Publique France

IA, modélisation et outils numériques
au défi de la COVID

26 mai 2021

De StopCovid à TousAntiCovid Un an de roman feuilleton

Eric Fleury
Directeur du centre de recherche Inria de Paris
Product Owner de l'application TousAntiCovid

Cette intervention est faite en toute indépendance vis-à-vis de l'organisateur de la manifestation. Je n'ai pas de conflit d'intérêts en lien avec le sujet traité.



Mars 2020

· Réflexion sur le contact tracing au sein de CARE (Comité d'Analyse, Recherche et Expertise)

8 Avril 2020

· GO officiel PR

Mai 2020

· Vote à l'assemblée Nationale et décret au conseil d'état

2 Juin 2020
12:00

· 🚀 Sortie dans les stores

Septembre 2020

· TAC : Relooking et nouvelles fonctionnalités

Avril 2021

· Passe sanitaire

Mai 2021

· Contact Tracing de lieu

Première fois que l'on vote à l'assemblée nationale au sujet d'une application



Mars 2020

· Réflexion sur le contact tracing au sein de CARE (Comité d'Analyse, Recherche et Expertise)

8 Avril 2020

· GO officiel PR

Mai 2020

· Vote à l'assemblée Nationale et décret au conseil d'état

2 Juin 2020 12:00

· 🚀 Sortie dans les stores

Septembre 2020

· TAC : Relooking et nouvelles fonctionnalités

Avril 2021

· Passe sanitaire

Mai 2021

· Contact Tracing de lieu

Première fois que l'on vote à l'assemblée nationale au sujet d'une application

8 avril → 2 juin
55 jours pour faire à la fois la recherche, sa conception et sa mise en production



Mars 2020

· Réflexion sur le contact tracing au sein de CARE (Comité d'Analyse, Recherche et Expertise)

8 Avril 2020

· GO officiel PR

Mai 2020

· Vote à l'assemblée Nationale et décret au conseil d'état

2 Juin 2020 12:00

· 🚀 Sortie dans les stores

Septembre 2020

· TAC : Relooking et nouvelles fonctionnalités

Avril 2021

· Passe sanitaire

Mai 2021

· Contact Tracing de lieu

Première fois que l'on vote à l'assemblée nationale au sujet d'une application

Fail Fast
Application taïseuse à :
- Application beaucoup plus interactive
- fonctionnalités qui donnent le sens d'une application mobile

Postulat 1 : un outil pour **2020/2021**
pour un état **démocratique** tel que
la France

Hypothèses

- Des institutions de confiance et une agence de protection des données indépendante de confiance
- Des utilisateurs curieux voire malveillants, et des risques de discrimination

Postulat 2 : en avril 2020, ne rien faire n'était pas envisageable

- Car « ça pourrait marcher », on le saura à **posteriori**... Aucun responsable politique n'aurait pris le risque de ne rien proposer
- Car il vaut mieux **être préparé** et avoir une solution qui ne sert pas, que de subir car on n'a rien anticipé
- Car nous (Inria) étions convaincus qu'une approche **équilibrée** était possible

Axiome: Privacy by design

- La donnée médicale personnelle « je suis positif » ne quitte pas le téléphone, ne circule pas, n'est pas stockée

« [...] nous pensons que le graphe social est bien plus sensible que le statut infecté »

Giles Hogben (Director of Privacy Engineering, Android/Google), 2 octobre 2020

https://www.youtube.com/watch?v=0gqZJX009Ko&feature=emb_logo [offset 3:35]

- Audit de la CNIL et code en open source
- Souveraineté pour l'autorité de santé



Efficacité



Souveraineté



Respect de la
vie privée



Démarche pour le développement de TAC

Transparence maximum en publiant

- les spécifications
- le code source de l'application et du serveur



Implication immédiate de l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI)

- les experts sécurité de la France
- expertise clef pour les spécifications, développement, opérationnel



Implication immédiate de la CNIL

- s'ils ne sont pas d'accord, ils peuvent tout bloquer !



Contact tracing : de quoi parle-t'on ?

Avertir toute personne qui a été *suffisamment proche* et *suffisamment longtemps* d'une ou plusieurs personnes testées COVID+

Un outil **complémentaire** au traçage manuel :

- entièrement automatique → passage à l'échelle
- anonyme pour ses utilisateurs → pas de risque de discrimination

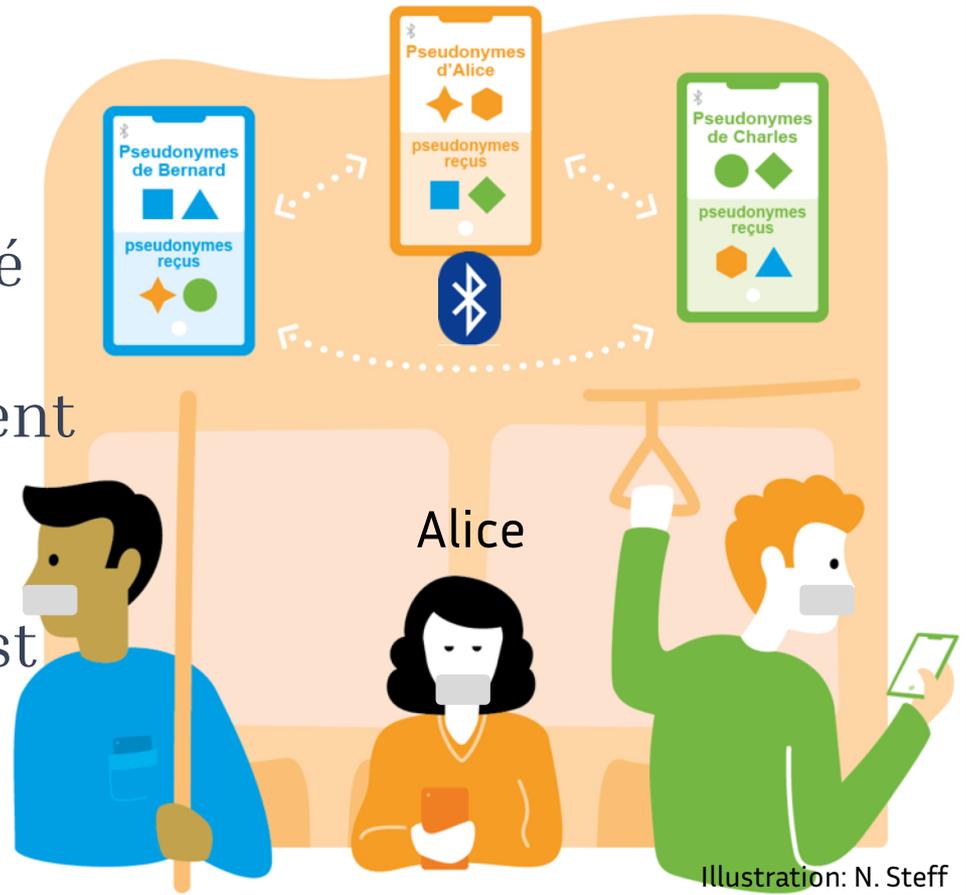
4 CONCLUSION

- 48 The world is facing a significant public health crisis that requires strong responses, which will have an impact beyond this emergency. Automated data processing and digital technologies can be key components in the fight against COVID-19. However, one should be wary of the “ratchet effect”. It is our responsibility to ensure that every measure taken in these extraordinary circumstances are necessary, limited in time, of minimal extent and subject to periodic and genuine review as well as to scientific evaluation.
- 49 The EDPB underlines that one should not have to choose between an efficient response to the current crisis and the protection of our fundamental rights: we can achieve both, and moreover data protection principles can play a very important role in the fight against the virus. European data protection law allows for the responsible use of personal data for health management purposes, while also ensuring that individual rights and freedoms are not eroded in the process.

Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak Adopted on 21 April 2020

1- Chaque application diffuse un pseudonyme via le Bluetooth...

- et enregistre les pseudonymes des smartphones à proximité
- les pseudonymes changent toutes les 15 minutes
- la portée du Bluetooth est limitée, heureusement



<https://hal.inria.fr/hal-02611265/document>

2- Alice, testée COVID+, remonte son historique au serveur...

- c'est l'ensemble des pseudonymes collectés dans les jours qui précèdent
- totalement anonyme (l'application d'Alice n'est pas identifiée)
- le serveur calcule le risque pour chaque contact

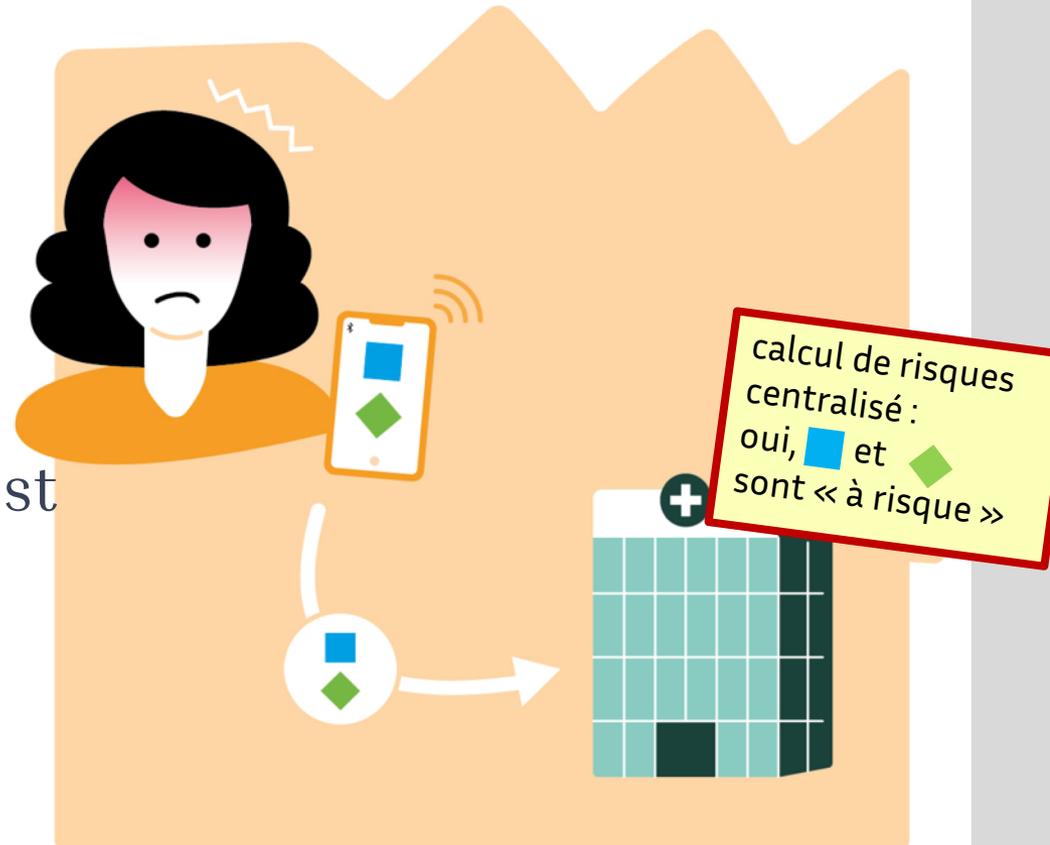


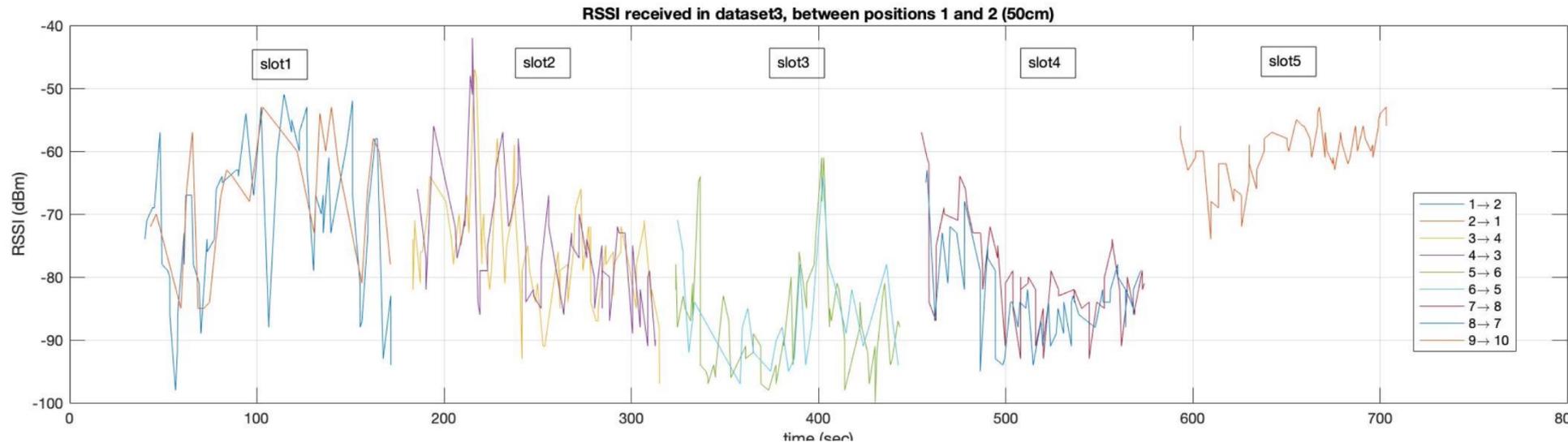
Illustration: N. Steff

3- L'appli de Bob interroge le serveur périodiquement

- ... le serveur l'informe qu'il a été suffisamment proche, suffisamment longtemps, avec une ou plusieurs personnes testées COVID+



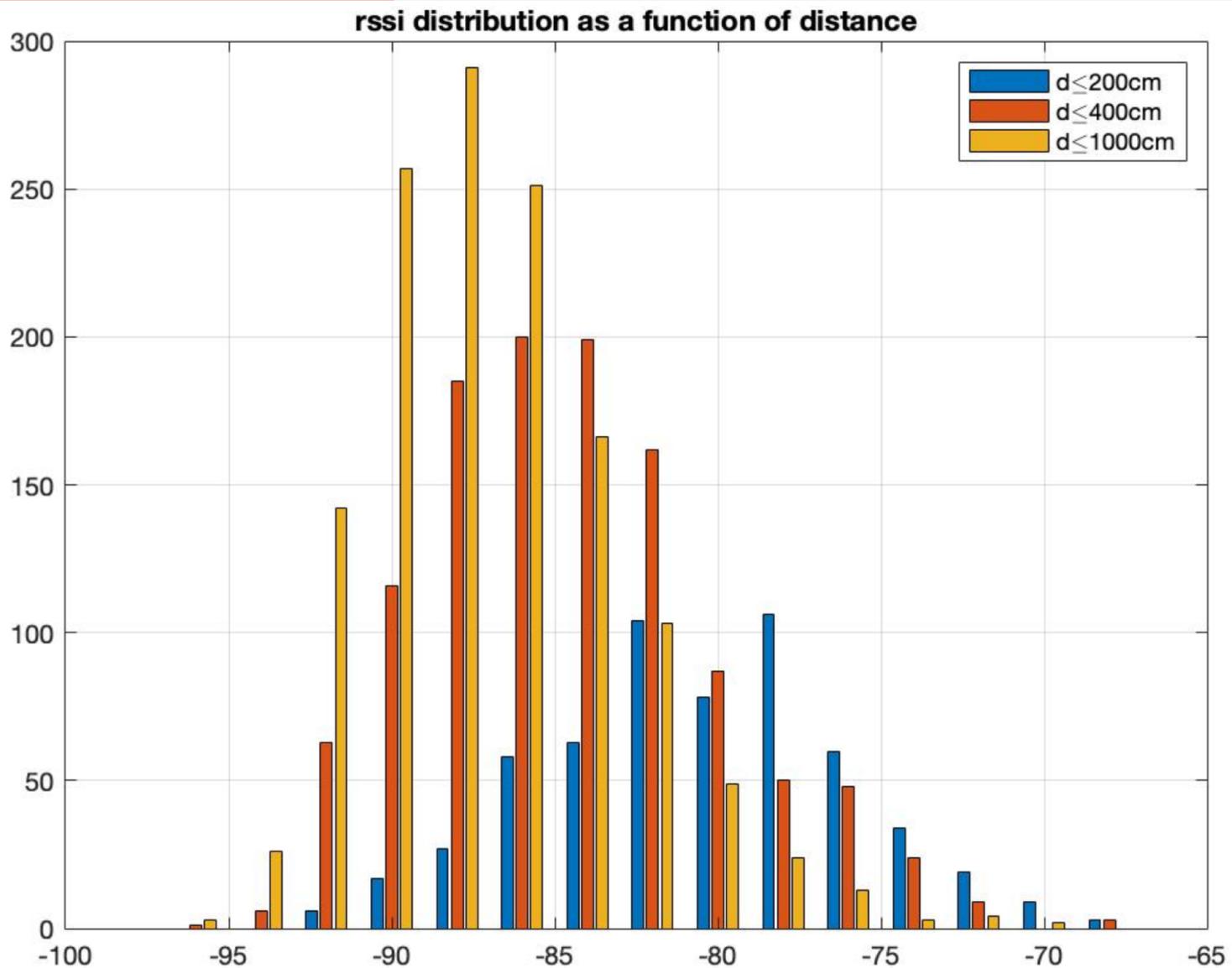
C'est quoi *suffisamment* proche pour du BLE?



Variabilité de la mesure BLE-RSSI.

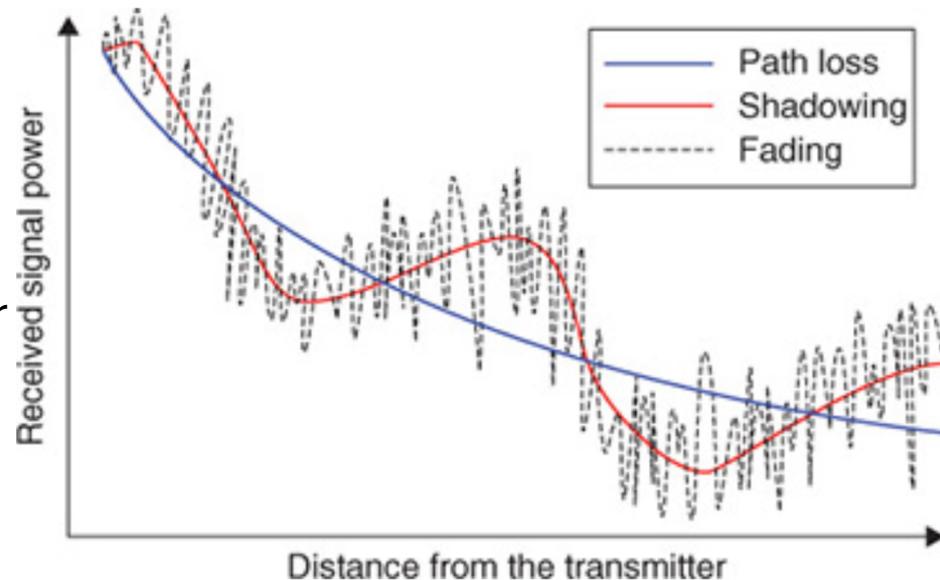
Illusoire d'avoir une bijection distance / RSSI par une simple moyenne

<https://hal.inria.fr/hal-02641630/document>



Phénomènes à prendre en compte

- **Shadowing ou masquage** : induit par la présence de personnes dans l'environnement. Même à 1m le positionnement du téléphone induit ce genre de masquage
- **Erreurs statiques** : composant émetteur et récepteurs.
- **Fading ou évanouissement** : nombreux chemins multiples.



- **Shadowing** : 5db
- **Erreurs statiques** : calibration
- **Fading ou évanouissement** : moyenne glissante

$$\hat{P}_{ji} = 10 \cdot \log_{10} \left(\frac{1}{m} \sum_{n=1}^m 10^{P_{ji}(n)/10} \right)$$

Calcul d'un score pour chaque contact

- Score d'exposition : distance + modèle de propagation
 - > On ajuste la confiance en fonction du nombre de mesure (Delta)
- Moyenne softmax pour le score moyen :
 - > une valeur élevée de RSSI est un témoin de proximité

$$\pi_{ij}[k] = \frac{RSSI[k] - P_0}{\Delta(N[k])}$$

$$\bar{\pi}_{jId}^{sm} = b \cdot \log \left(\frac{1}{N} \sum_{n=1}^N e^{\pi_{jId}(n)/b} \right)$$

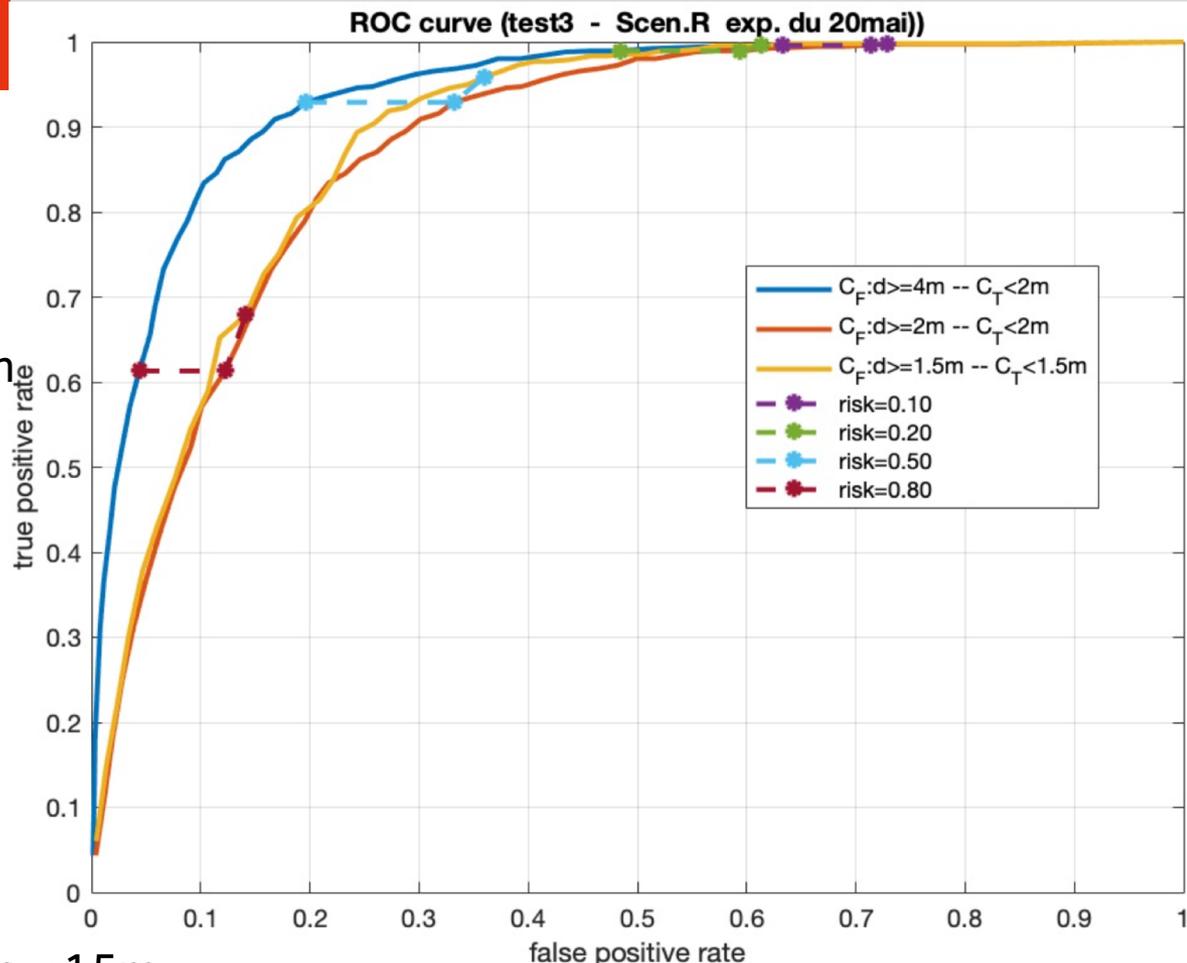
Calcul du risque d'exposition

- Risque de contamination d'un individu sachant le score de ses contacts :

$$\mathbb{P}(\text{IS}_j = I | \Pi(j)) = 1 - \exp \left(-r_0 \cdot \sum_{\ell=1}^L \pi_{\ell} \right)$$

En pratique...

- Résultat de score vu comme une fonction de classification
 - > Je suis proche < 1.5m
 - > Entre 2m et 4m
 - > Eloigné > 4m
- Area Under Curve > 90%
- Seuil à 0.8 (rouge)
 - > On garde 70% des contacts < 1.5m et 60% < 2m
 - > On rejette 88% > 2m et 96% > 4m



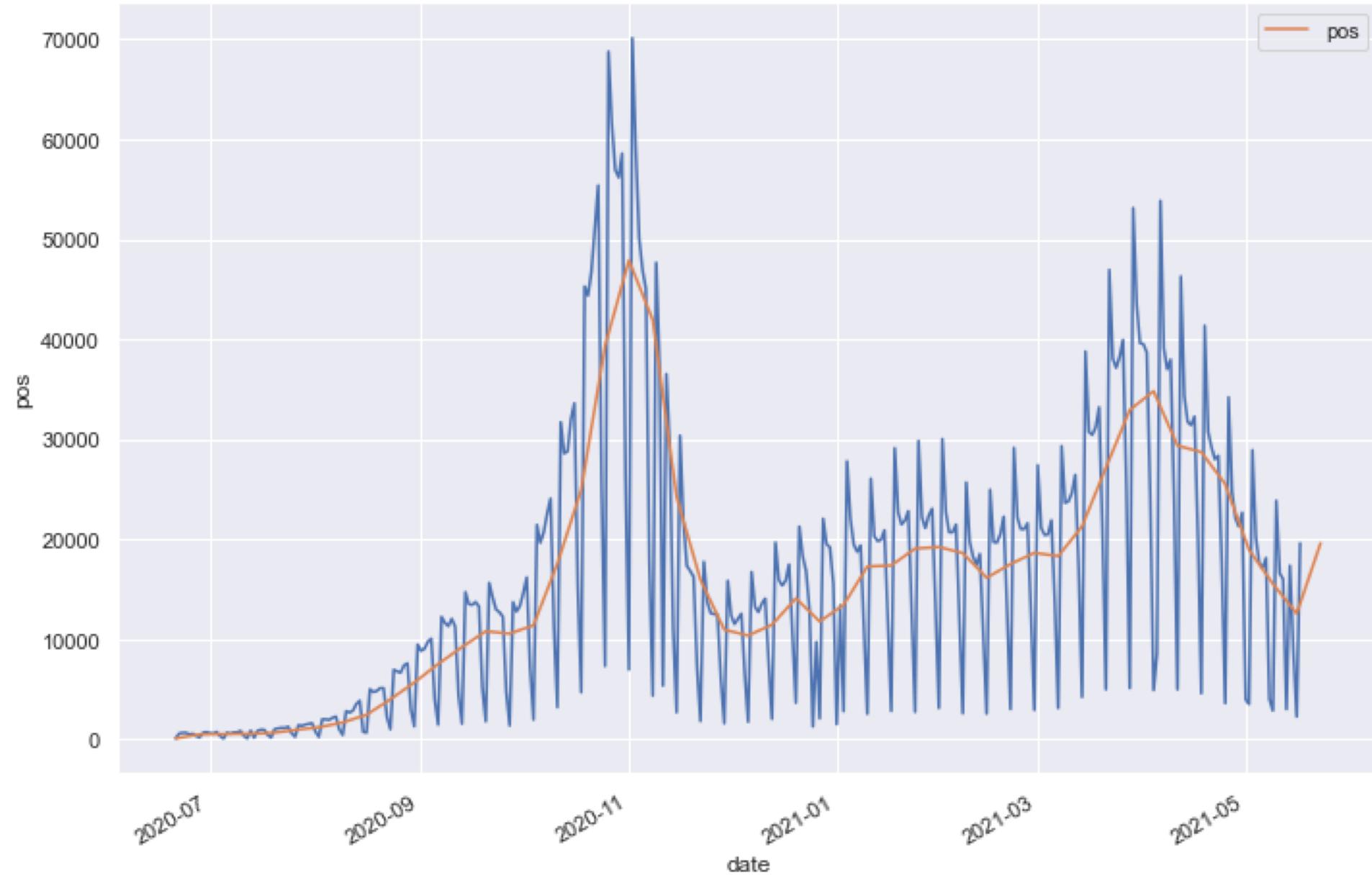
Être « centralisé » pour une efficacité maximum

→ l'analyse de risque est opérée par un serveur central qui détient des informations sur les applications enregistrées

Être centralisé est une nécessité pour une meilleure efficacité épidémiologique

- connaître en temps réel :
 - le nombre de remontées de personnes COVID+
 - le nombre de notifications « à risque »
- contrôler tous les paramètres du calcul de risques
 - par ex. à ajuster à la capacité de tests, suivre la connaissance scientifique
- le MSS doit pouvoir piloter
 - un aspect clé face à une crise inédite

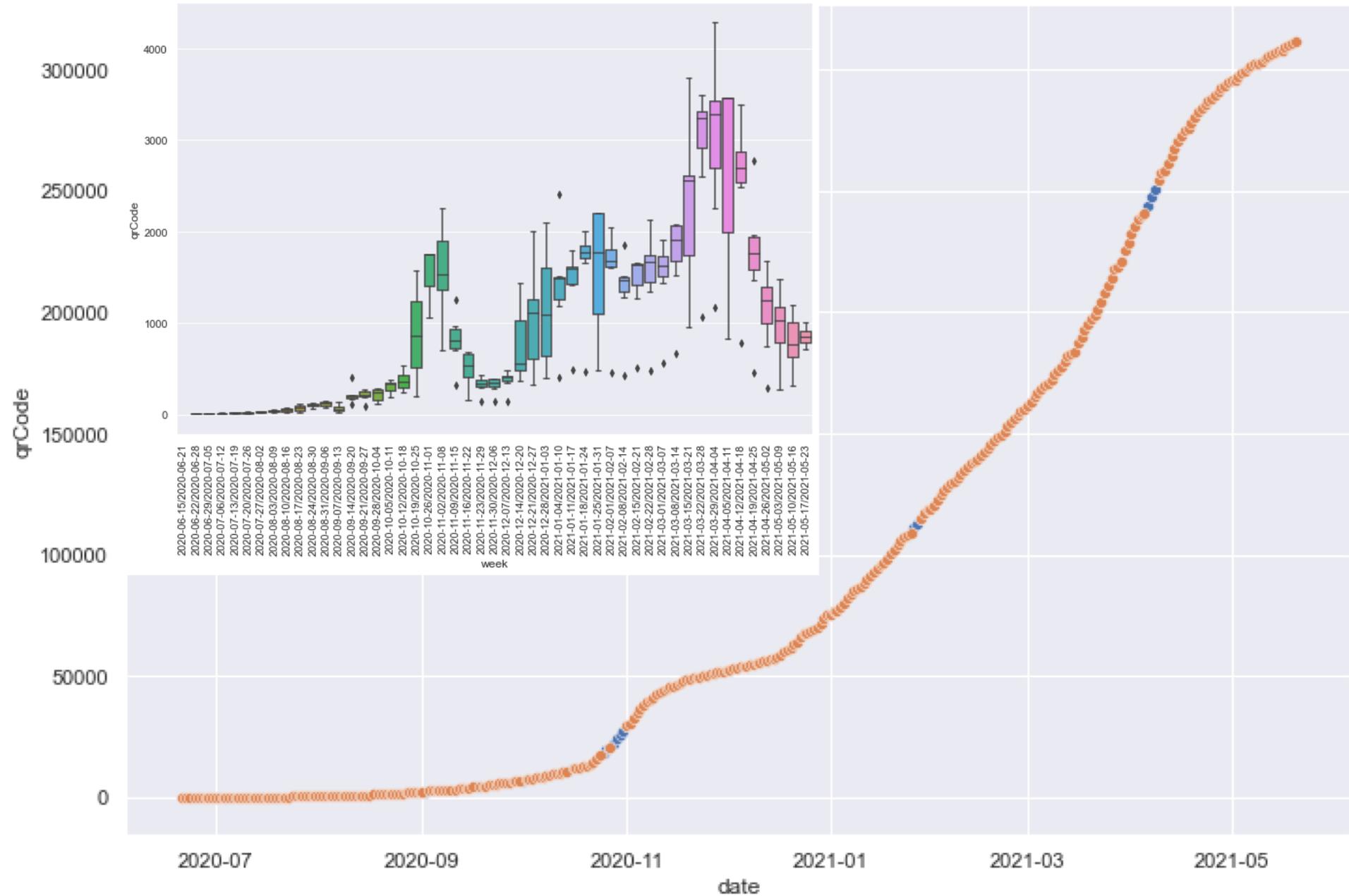
Nombre de cas positifs en France



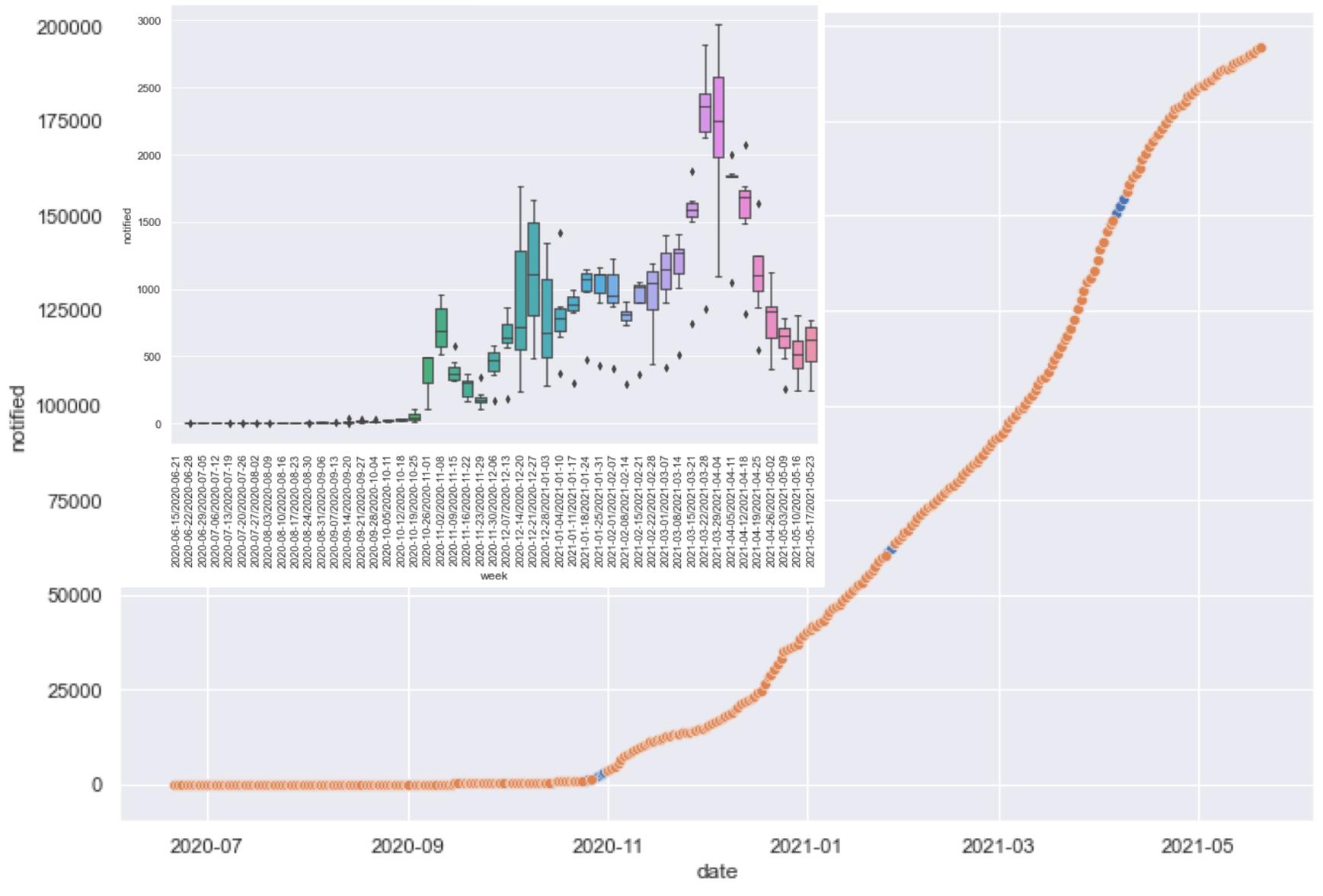
Nombre d'applications enregistrées



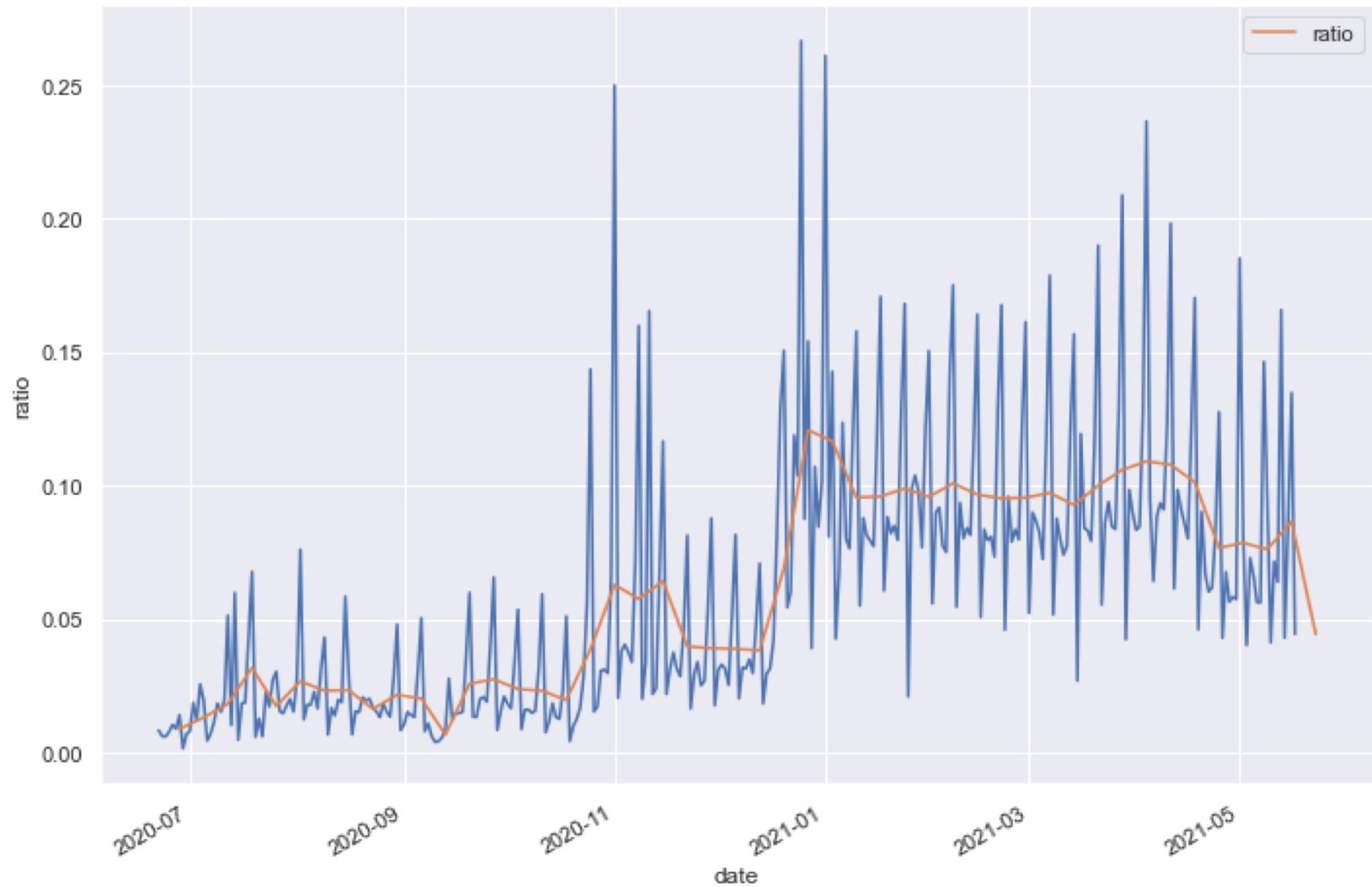
Nombre de cas déclarés dans TousAntiCovid

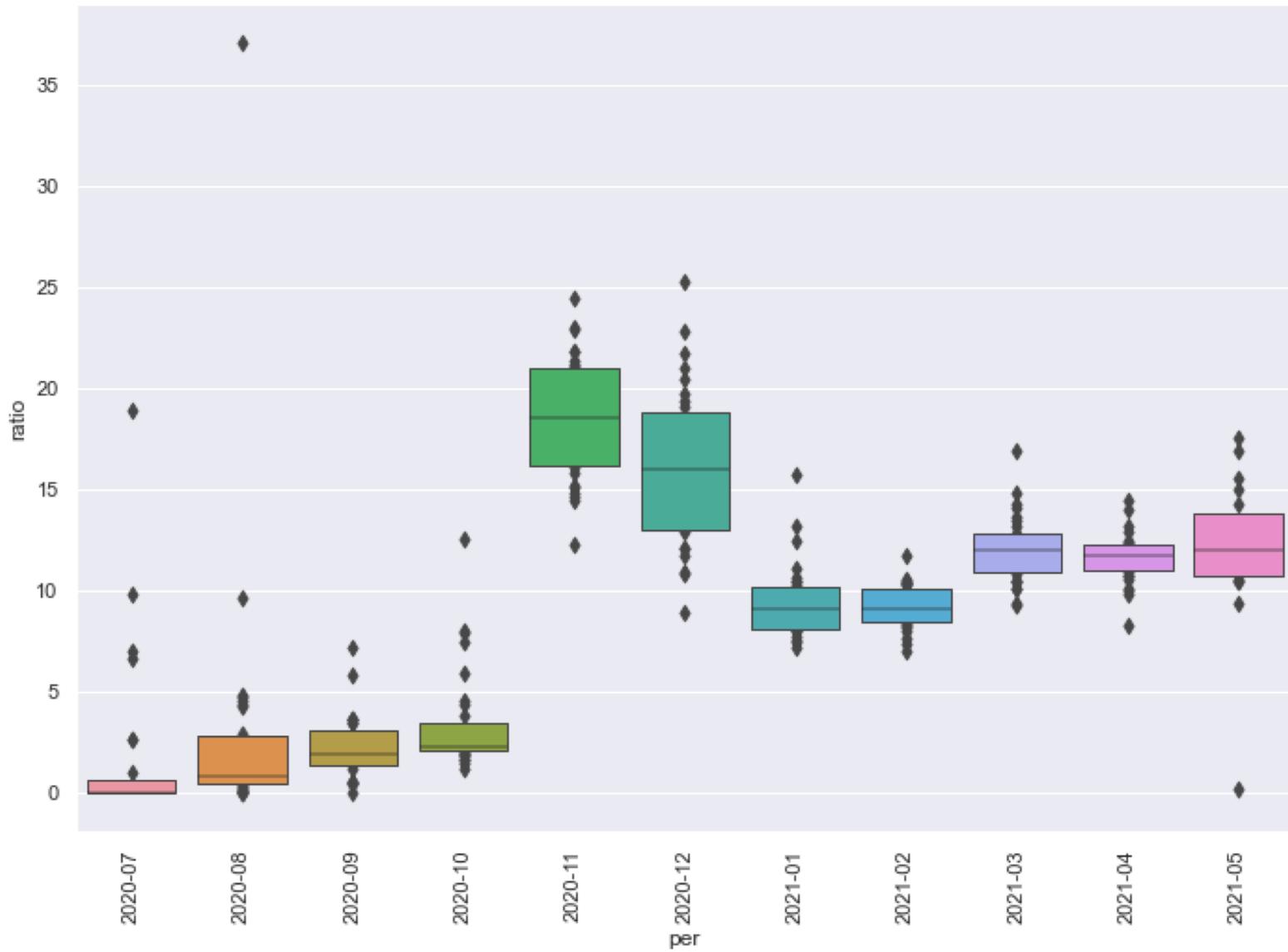


Nombre de notifications envoyées par TAC



Ratio Nombre de notifications / Nombre de cas positifs





Contact tracing de lieu

Cibler les lieux clos où les gestes barrières sont difficilement respectables



Détecter automatiquement les lieux à risque (Contact tracing)

Détecter automatiquement les lieux potentiellement cluster (Backward Tracing)

Avertir toute personnes présentes dans un lieu jugé à risque sur la même plage horaire que la ou les personnes qui se sont déclarés positives

1- Alice scanne un QR code à l'entrée d'un lieu

le QR code est stocké sur le téléphone d'Alice (et y reste)

contient notamment un pseudonyme du lieu + horodatage lors du scan

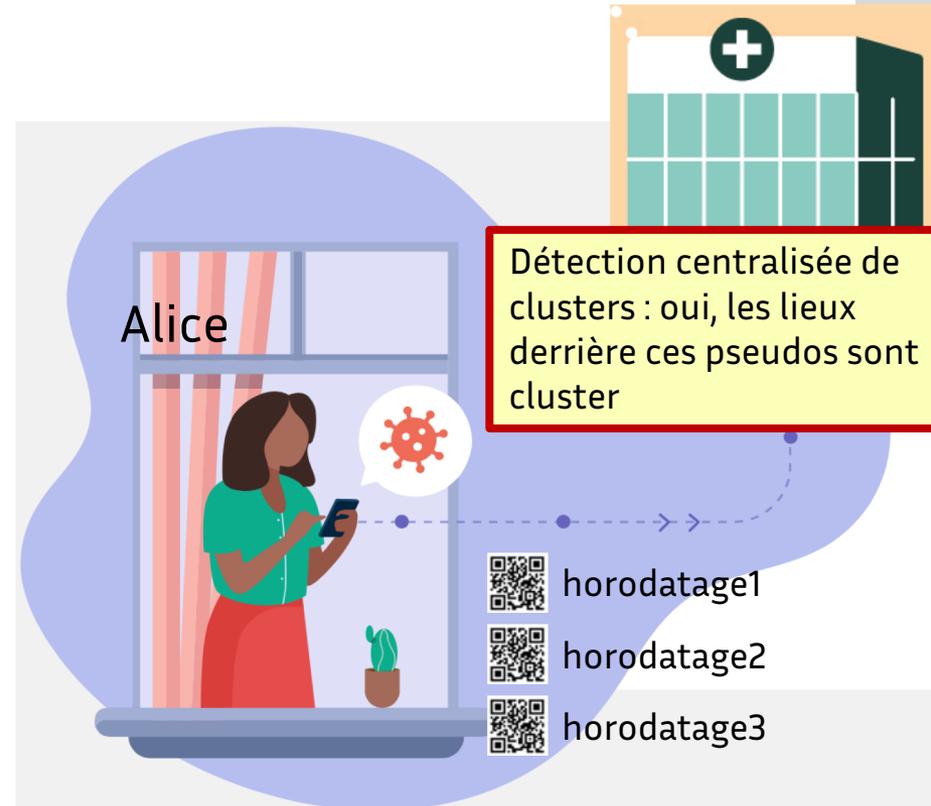


2- Alice, testée COVID+, remonte son historique au serveur...

si Alice y consent

traitement anonyme sur le serveur
(l'application d'Alice n'est pas
identifiée)

le serveur calcule le risque et met à jour
une « liste noire » des pseudonymes
des lieux cluster



3- Les applis vérifient périodiquement le risque d'exposition localement

les applications récupèrent la liste noire des lieux cluster

chaque application vérifie si l'utilisateur s'est rendu dans un lieu cluster en comparant avec l'historique de QR codes scannés

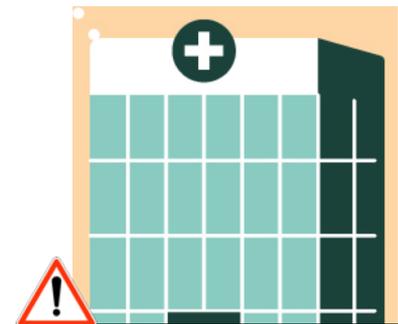


Pourquoi une vérification du risque décentralisée ?

vie privée : car la majorité des utilisateurs ne partageront jamais les pseudonymes des lieux visités avec quiconque

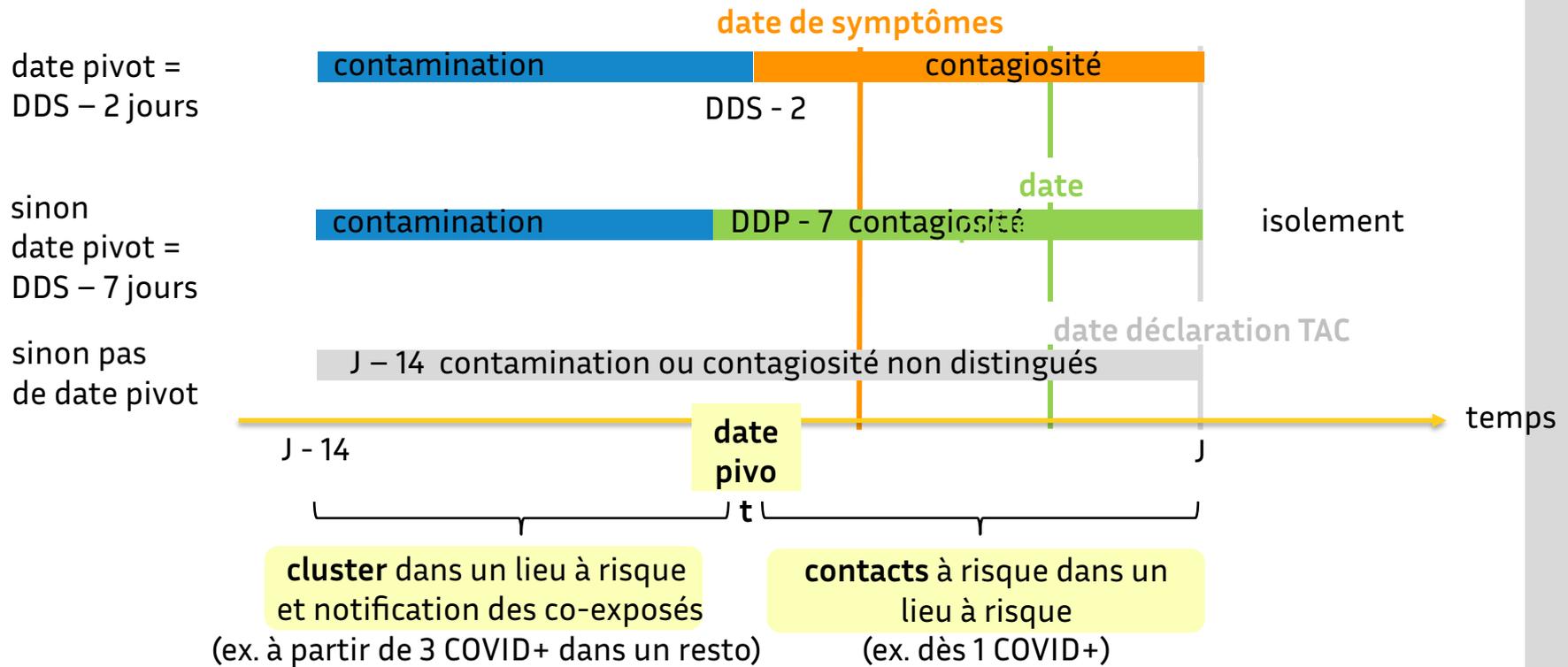
- peuvent être porteur de sens
- même si tout est fait pour éviter des dérives sur le serveur, même si ce ne sont que des pseudonymes

permis par le RGPD : la liste noire ne contient que des pseudonymes de lieux cluster, ce ne sont pas des données de santé



liste noire des {pseudos +
tranche horaire} des lieux cluster

Permet d'identifier des clusters



TousAntiCovid est une application à succès, à minima pour ses fonctions pratiques et d'information

L'efficacité de TAC sur les fonctions de suivi de contacts et historique de lieux doit être évaluée

- demandée par la CNIL et nécessaire pour juger de la proportionnalité risques de vie privée / utilité
- pas encore été possible (manque de données / nouveau décret)

Tracing de lieu doit permettre d'identifier des clusters grâce à sa capacité « traçage arrière »

- des lieux ont-ils donné lieu à plusieurs contaminations ? Si oui → cluster
- pas possible en l'état avec le suivi de contacts à la ROBERT ou GAEN

Merci

