Synthèse du débat

Pause-Café Éthique et Data #3

17 juin 2020

Partagez votre expérience :
@adnouest

Arnauld CASTEX

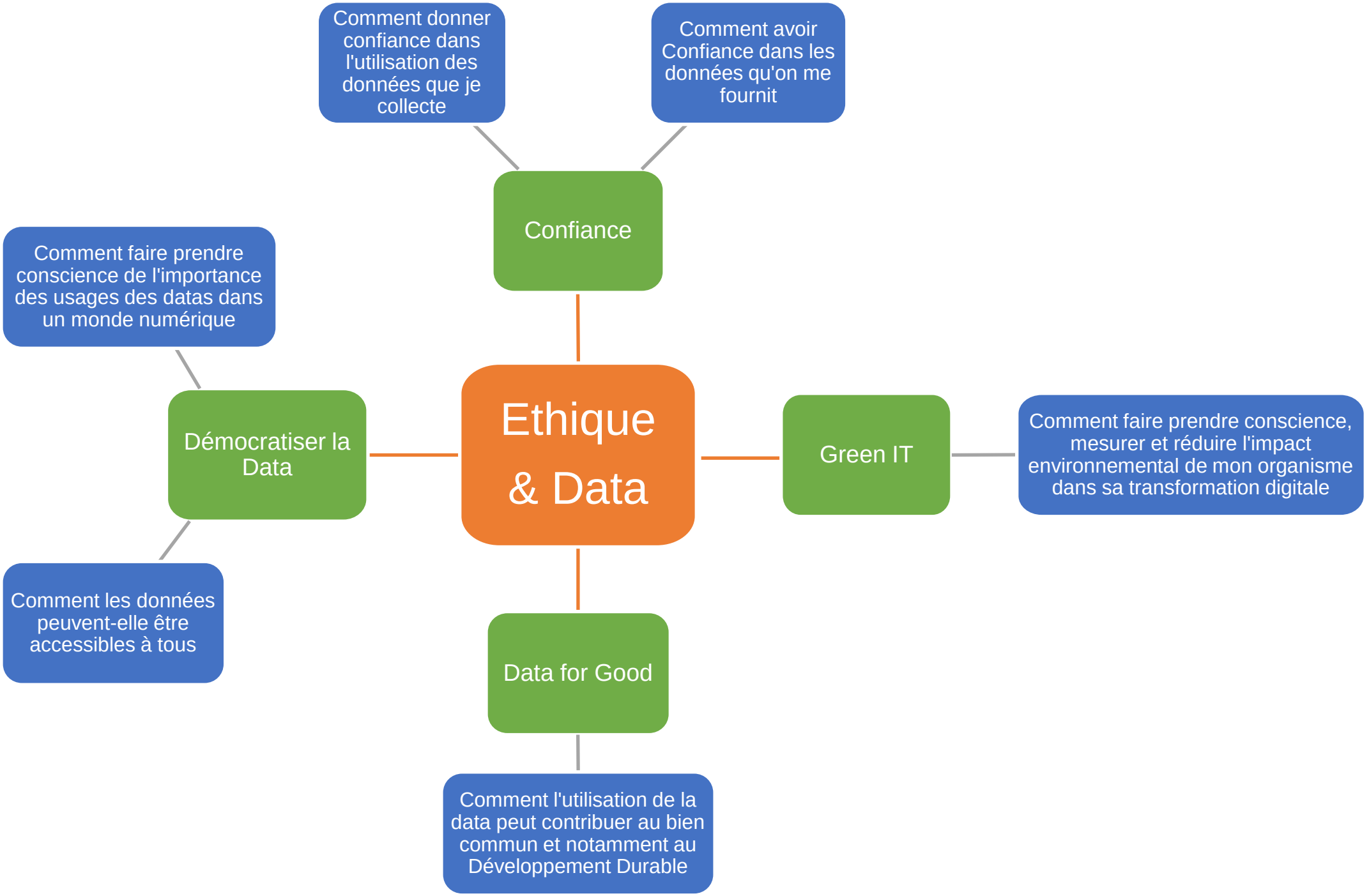
Christian BONNIN

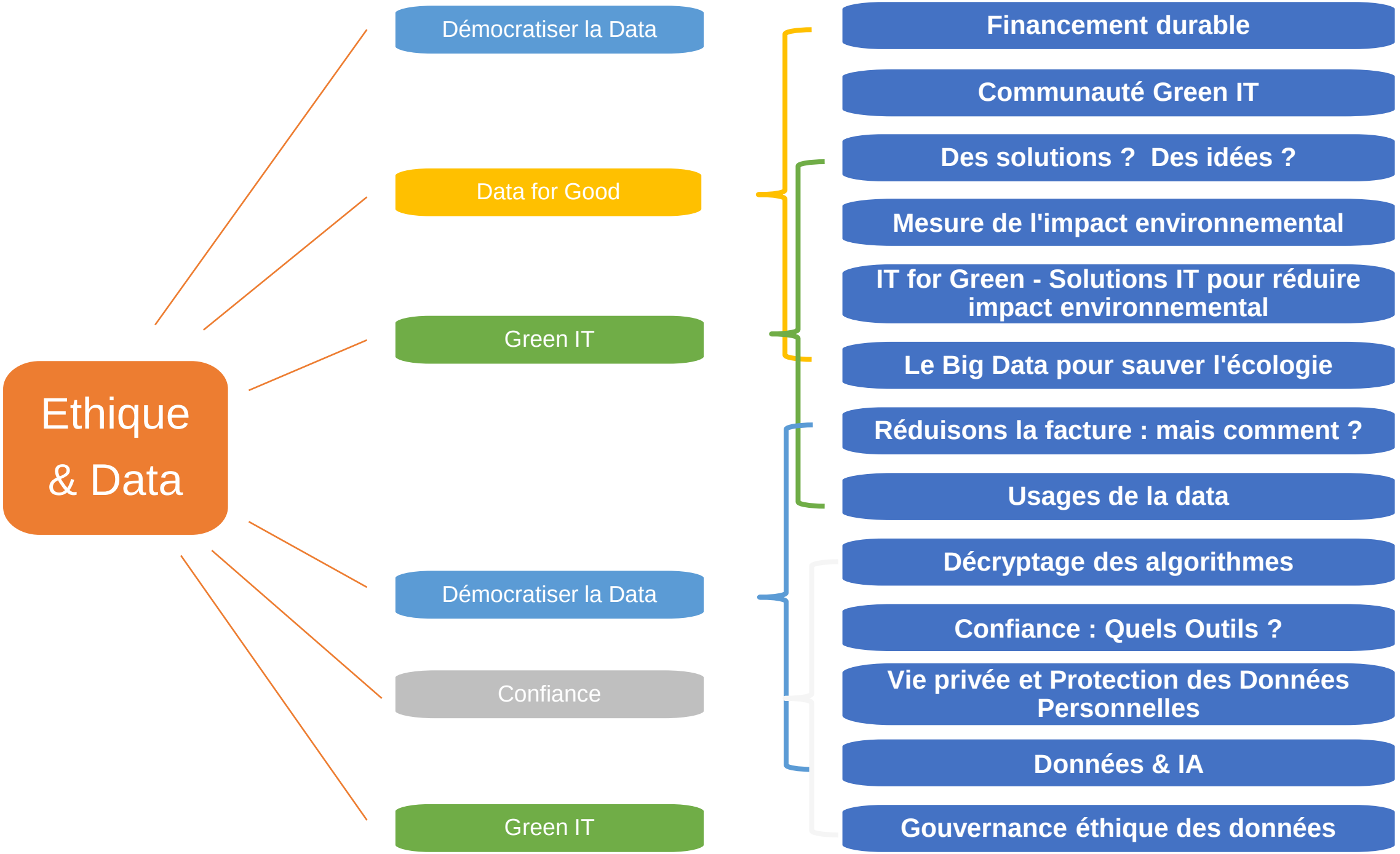
Elisabeth LEHAGRE

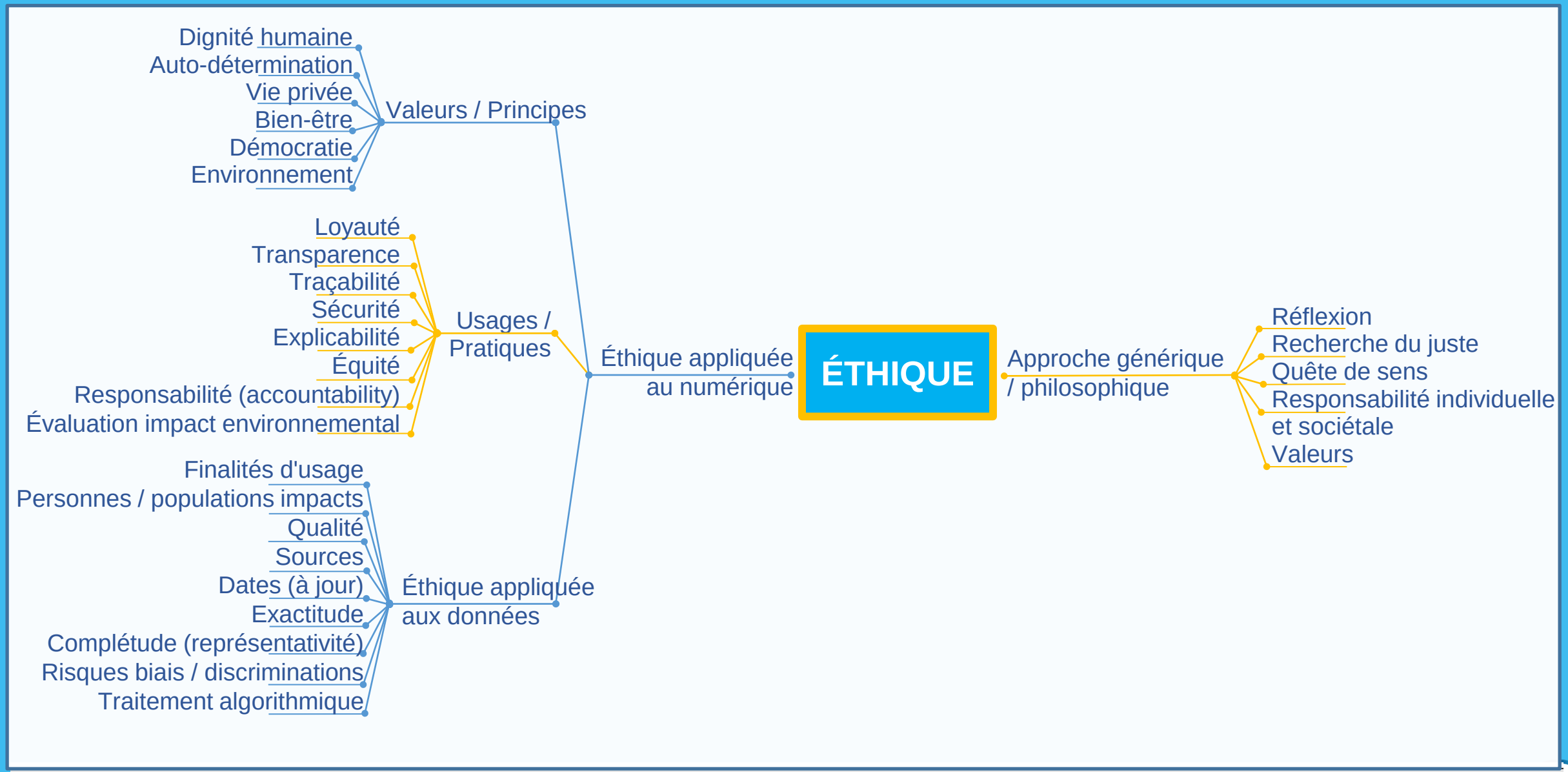
Rémi LE MAUFF

Christophe BURGUIN

Définitions

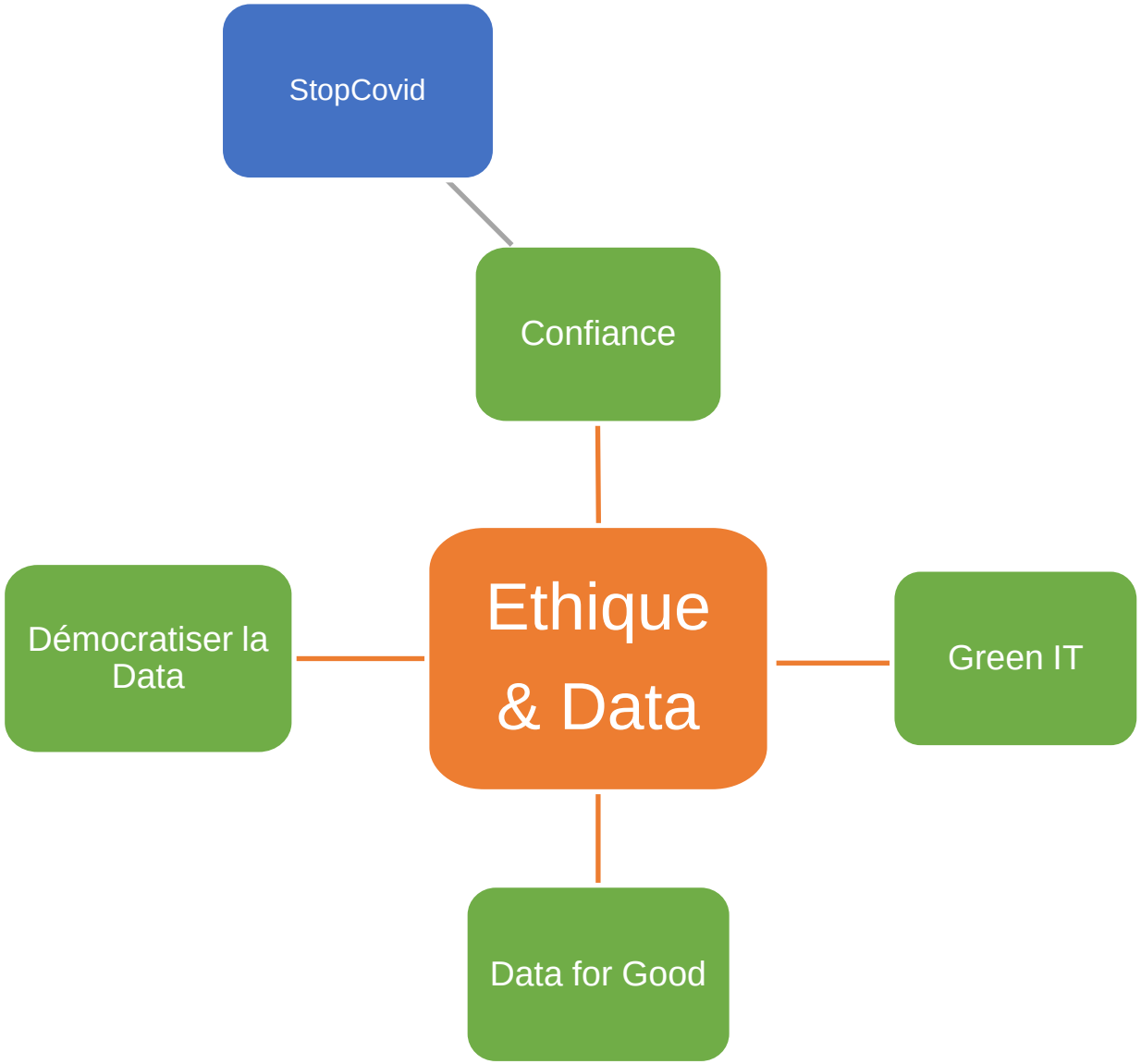



Ethique & Data
Démocratiser la Data
Data for Good
Green IT
Démocratiser la Data
Confiance
Green IT
Financement durable
Communauté Green IT
Des solutions ? Des idées ?
Mesure de l'impact environnemental
IT for Green - Solutions IT pour réduire impact environnemental
Le Big Data pour sauver l'écologie
Réduisons la facture : mais comment ?
Usages de la data
Décryptage des algorithmes
Confiance : Quels Outils ?
Vie privée et Protection des Données Personnelles
Données & IA
Gouvernance éthique des données



Rappel du contexte

Le sujet : StopCovid



Mounir MAHJOUBI, député de Paris

Tour d'horizon de l'usage des données de traçage

Traçage des données mobiles dans la lutte contre le Covid-19

De quoi parle-t-on?

L'usage du traçage des données mobiles dans la lutte contre la pandémie de Covid-19 répond à trois finalités :

1. L'observation des pratiques collectives de mobilité et de confinement (i.e. cartographie des déplacements de population).
2. L'identification des sujets "contact" (i.e. backtracking ou contact tracking).
3. Le contrôle des confinements individuels (i.e. tracking ou bracelet électronique virtuel).

Plusieurs technologies supportent ces usages à travers le monde :

1. le bornage téléphonique,
2. des applications GPS,
3. des applications Bluetooth,
4. les systèmes de cartes bancaires et de transport,
5. la vidéosurveillance, dotée ou non d'intelligence artificielle.

Résumé de la NOTE PARLEMENTAIRE de Mounir MAHJOUBI, député de Paris

(Version 1.0 du lundi 6 avril 2020)

2 usages moins concernés par le débat

Observation des pratiques collectives de mobilité et de confinement

- *Finalités :*
  - *Obtenir une vision nationale et régionale*
  - *Obtenir une vision affinée à l'échelle d'un quartier*
- Analyses déjà réalisées aujourd'hui au niveau macro avec un bon niveau de confidentialité
  - France - Partenariat Orange Inserm : Observer les pratiques collectives de mobilité
  - Italie - Observer les pratiques collectives de confinement à l'échelle d'une ville
- Le focus sur des zone géographiques réduites peut se révéler sensible, en stigmatisant des populations
  - ¼ des parisiens a quitté Paris avant le confinement

Le contrôle des confinements individuels

- *Finalités :*
  - *Veiller au respect des quarantaines (sujets malades et "contact")*
  - *Veiller au respect des confinements (population générale)*
  - *Développer un "permis de circuler"*
- *Se rapproche d'une détention à domicile de type bracelet électronique :*
  - Taïwan (appel inopiné)
  - Pologne (avec *selfie*)
  - Hong Kong (bracelet, sinon appel inopiné)
- En contradiction avec de nombreuses valeurs des pays européens.

L'usage débattu : Identification des sujets “contact”

Finalités :

- *Retracer le parcours récent des personnes testées positives*
- *Informer la population des zones à risque*
- *Relever directement les contacts récents entre les individus testés positifs et des personnes tierces*

Pour éviter un rebond post-confinement, les experts recommandent d’identifier les personnes ayant récemment été au contact d’un sujet infecté.

L’enjeu est de leur proposer un test de dépistage et d’opérer leur mise en quarantaine. Pour cela, il est nécessaire de retracer le parcours récent des patients testés positifs.

Identification des sujets “contact” (1/2)

Données issues du bornage des opérateurs télécoms

- *Fonctionnement : utilisation des données des antennes télécom.*
- *Exemples*
  - TAIWAN - Un itinéraire rendu public
  - CORÉE DU SUD - La Corée du Sud rend publics les déplacements récents des personnes testées positives au Covid-19. Ils sont à l’origine de nombreuses applications qui informent leurs utilisateurs.
- Enjeux éthiques
  - Manipulation de données personnelles nominatives => nécessite un consentement
  - Création d'une base de données => nécessite un cadre :
    - Durée de rétention
    - Durée de l'application
    - Usage des données dans un seul but

Données GPS issues d’applications mobiles

- Trois usages :
  1. Retracer le parcours récent des personnes testées positives
  2. Informer la population des zones à risque
  3. Relever directement les contacts récents entre les individus testés positifs et des personnes tierces
- *Exemples*
  - CHINE - Close Contact Detector : une app’ en libre téléchargement, opaque et reliée aux bases de données gouvernementales
  - MIT - 2 objectifs : stocker des informations fiables sur le parcours des 28 derniers jours et informer les personnes “contact” tout en protégeant les données personnelles
- Enjeux éthiques
  - Idem bornage si les données sont envoyées à un serveur
  - Possibilité de crypter les informations transmises, mais nécessite le consentement
  - Si info de risque descendantes, alors pas de transfert de données personnelles
- Limite : Efficace si l'application est utilisée par 60% de la population et si elle est accompagnée d'une campagne de dépistage.

Identification des sujets “contact” (2/2)

Connexions Bluetooth issues d’applications mobiles

- *Fonctionnement :*
  - Un historique de tous les téléphones ayant été en contact avec la personne est enregistré dans le téléphone.
  - Lorsqu’un utilisateur est testé positif, il se déclare dans l’application et tous les utilisateurs tiers qui ont été à son contact sont informés en recevant une alerte sans qu’il y ait eu besoin de déclarer leur identité. .
- *Exemple*
  - SINGAPOUR - TraceTogether : Une application gouvernementale en libre téléchargement
- Enjeux éthiques
  - Idem données GPS

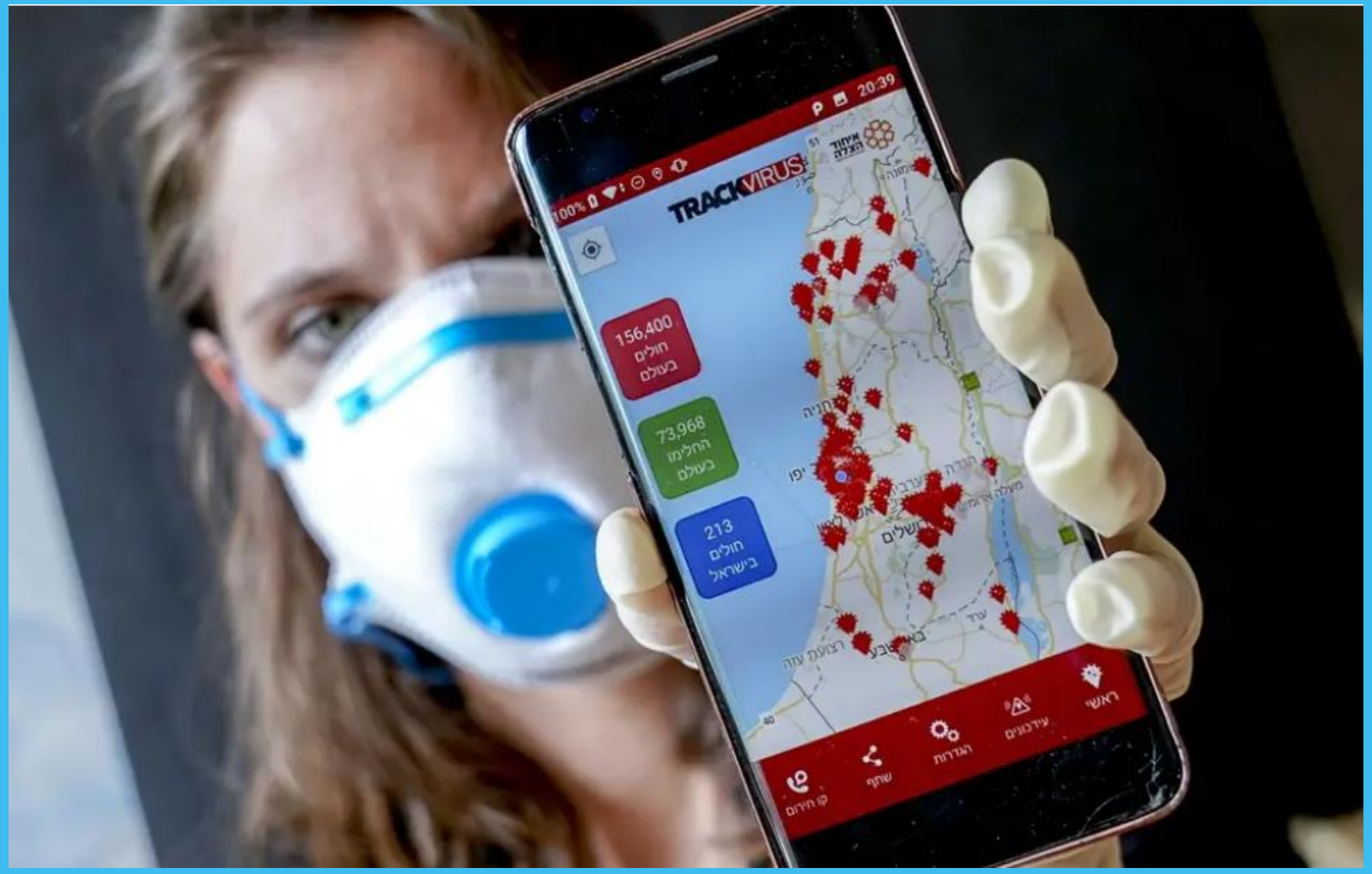
Données issues des cartes bancaires et de transport

- Usage :
  - Lorsque le personnel d’un point de vente (/usager des transports) est testé positif, la méthode permet d’identifier les clients “*contact*” (/usagers “*contact*”).
  - Les transports en commun sont des espaces à risque élevé (étroits, très fréquentés, position statique des personnes durant plusieurs minutes). Il est donc intéressant de pouvoir y observer les déplacements et les proximités entre usagers.
- *Exemple*
  - SINGAPOUR, CORÉE DU SUD - L’usage permet d’identifier des boutiques, des transports et des stations à risque
- Enjeux éthiques
  - Obtention du consentement des personnes avant le *backtracking*
  - Une information des sujets “*contact*” mais non des autorités sanitaires
  - Une gouvernance de contrôle des usages.

L'utilisation de la vidéosurveillance n'a pas été retenue car trop éloignée des valeurs démocratiques (Big Brother).

Identification des sujets “contact” - synthèse

	CHI	TAÏ	COR	SIN	ISR	RUS	POL	USA	ITA	ALL	FRA
Usages											
1) Observer les pratiques collectives											
- Obtenir une vision nationale et régionale	X	X	X	X	X	X	?	X	X	X	X
- Obtenir une vision affinée (i.e. quartiers)	?	?	?	?	?	?	?				
2) Identifier des sujets *“contact”*											
- Retracer le parcours récent des “positifs”	X	X	X	X	X					(4)	
- Informer la population des zones à risque		X	X								
- Relever directement les contacts récents				X				(5)		(5)	(5)
3) Contrôler des confinements individuels											
- Veiller aux quarantaines (malades+contact)	X	X	X		X	X	X				
- Veiller aux confinements (pop. générale)		(1)	(1)								
- Développer un *“permis de circuler”*	X										
Techniques											
Bornes téléphoniques	X	X	X	X	X	X	X	X	X	X	(3)
GPS	X	X	X	X	X	X					
Données bancaires / de transport	X	X	X	X		X					
Vidéo surv. + reconnaissance faciale	X	X	X	X		X					
Bluetooth				X						(5)	(5)



Sous le capot de l'application StopCovid

Les grands principes

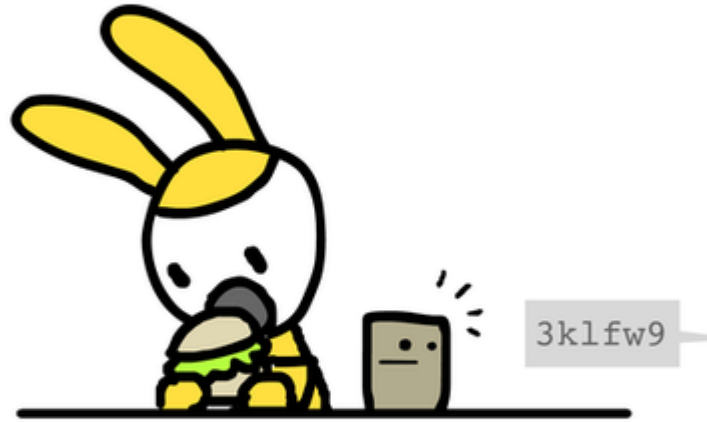
- Volontariat
- Anonyme
- Temporaire
- Notification si contact passé avec un malade
- Crypto Identifiants
- Pas de géolocalisation - Bluetooth
- Protocole technique ROBERT ou DP-3T

Alice télécharge une application de traçage (dont le code est public , ce qui fait qu'on peut vérifier que l'application fonctionne bien comme ceci :)

toutes les 5 minutes, son téléphone envoie un charabia aléatoire et unique à tous les autres téléphones à proximité, par Bluetooth.

* 5 minutes plus ou moins, c'est un exemple ! par ailleurs techniquement ce sont des messages peudo-aléatoires mais ça ne change rien à notre affaire.

Dans la mesure où les messages sont aléatoires et ne contiennent aucune donnée de géolocalisation, ils ne contiennent AUCUNE information sur l'identité d'Alice ni sur les endroits où elle est allée.



En plus d'envoyer ses messages aléatoires, le téléphone d'Alice reçoit également les messages des autres téléphones à proximité.

Par exemple, celui de Bob.

Bob a installé la même application de traçage sur son téléphone (ou une autre application de traçage respectueuse de sa vie privée et compatible avec celle d'Alice).

Si Bob et Alice restent à proximité plus de 5 minutes, leurs téléphones vont échanger un message aléatoire et unique.

Leurs téléphones se souviennent de tous les messages dits et entendus ces derniers 14 jours.

Rappel: les messages aléatoires ne contiennent AUCUNE INFO, l'intimité d'Alice est protégée de Bob, et vice versa!

* 14 jours est aussi juste un exemple! les épidémiologistes pourraient apprendre que la "période infectieuse" est en fait plus courte ou plus longue que cela.

Le lendemain, Alice développe une toux sèche et de la fièvre.

Alice se fait tester.

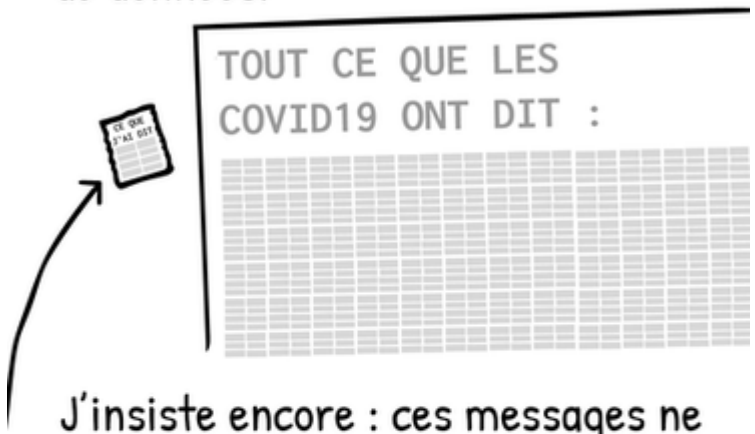
Alice a le COVID-19.

Ce n'est pas un bon jour pour Alice.

Mais elle peut maintenant se venger du virus ! Alice envoie la liste des messages qu'elle a envoyés ces 14 derniers jours à un hopital, en utilisant un code qui lui a été fourni à cet effet par son médecin (c'est pour éviter le spam).

Alice peut également cacher les messages qu'elle veut garder pour elle, comme ceux que l'application a envoyés quand elle était à la maison.

L'Hôpital range les messages aléatoires d'Alice dans sa base de données.



J'insiste encore : ces messages ne contiennent aucune information sur les endroits où elle est allée, sur son identité, *ni même sur le nombre de personnes avec qui elle a été en contact* ! Ces informations ne sont d'aucun usage à l'hôpital.

* Différents hôpitaux peuvent mettre ces bases de données en commun, mais dans la mesure où aucune information personnelle n'y est stockée ça ne présente pas de danger.

...mais pas pour Bob !

Le téléphone de Bob vérifie régulièrement la liste de messages de cas du COVID-19 de l'hôpital, et voit s'il a "entendu" n'importe lequel de téléphones proches ces derniers 14 jours.
(Le charabia ne donne à Bob AUCUNE AUTRE INFO PERSONNELLE.)

* le vrai protocole DP-3T est encore PLUS sécurisé ! il utilise une méthode appelée "cuckoo filter" pour que les téléphones ne connaissent QUE les messages covid-19 qu'ils ont entendus, sans révéler TOUS les messages covid-19.

S'il a reçu, mettons, 6 messages envoyés par le téléphone de personnes atteintes de COVID19 (6 * 5 minutes : 30 minutes d'exposition au virus), le téléphone va suggérer à Bob de se placer en quarantaine.

Et ainsi Bob stoppe la propagation du virus, en étant un jour en avance sur lui !

* à nouveau, ces nombres ne sont que des exemples !

Les questions éthiques expliquées par l'exemple

La fausse déclaration

Le joueur de foot Gronaldo doit disputer le prochain match de Ligue des champions. Pour l'empêcher de jouer, il suffit pour un adversaire de laisser son téléphone à côté de celui de Gronaldo à son insu, puis de se déclarer malade. Gronaldo recevra une alerte, car il aurait été en contact avec une personne infectée, et devra rester 14 jours éloigné des terrains

Mes voisins sont-ils malades?

M. Ipokondriac voudrait savoir si ses voisins sont malades. Il récupère son vieux téléphone dans un placard, y installe l'application TraceVIRUS, et le laisse dans sa boîte aux lettres en bas de l'immeuble. Tous les voisins passent à côté à chaque fois qu'ils rentrent chez eux, et le téléphone recevra une notification si l'un d'entre eux est malade.

Suspect unique

M. Lambda qui, pour éviter la contamination, ne sort de chez lui que pour faire ses courses à l'épicerie du quartier, reçoit une notification de [illegible] onsable n'est autre que l'épicier.

Résumé

- Il n'y a pas de base de données nominative des malades. ☑ VRAI
- Les données sont anonymes. ⊘ FAUX
- Il est impossible de retrouver qui a contaminé qui. ⊘ FAUX
- Il est impossible de savoir si une personne précise est malade ou non. ⊘ FAUX
- Il est impossible de déclencher une fausse alerte. ⊘ FAUX
- L'utilisation du Bluetooth ne pose pas de problème de sécurité. ⊘ FAUX
- Ce dispositif rend impossible un fichage à grande échelle. ⊘ FAUX

Pour aller plus loin :

Pour aller plus loin, quelques liens utiles :

- *Le Github du protocole ROBERT :* https://github.com/ROBERT-proximity-tracing/documents
- *Le Github du protocole DP-3T :* https://github.com/DP-3T/documents
- *Framablog :* https://framablog.org/2020/04/12/une-appli-de-tracage-du-covid-9-qui-echappe-a-big-brother/

- *Les scenaris qui dérangent :* https://framablog.org/2020/04/24/applis-de-tracage-scenarios-pour-les-non-specialistes/
- *Avis d'experts :* https://risques-tracage.fr/docs/risques-tracage.pdf
- *L'avis de la CNIL :* https://information.tv5monde.com/info/application-stopcovid-et-libertes-la-cnil-pose-des-conditions-356882

Synthèse

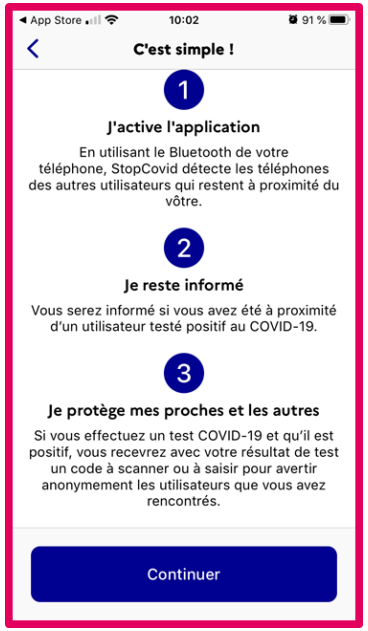
Application Stop Covid – éthique & data

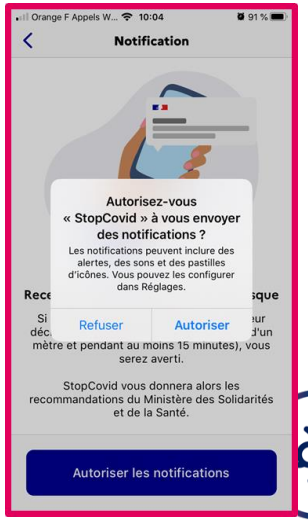
L'application Stop Covid est sortie les 2 et 3 juin sur les stores
Le groupe éthique et data vous invite à sa troisième pause-café afin de poursuivre les échanges sur l'utilisation des données face à la pandémie Covid-19.

2% de la population l'a téléchargé

350.000 utilisateurs actifs
(0,5% de la population)

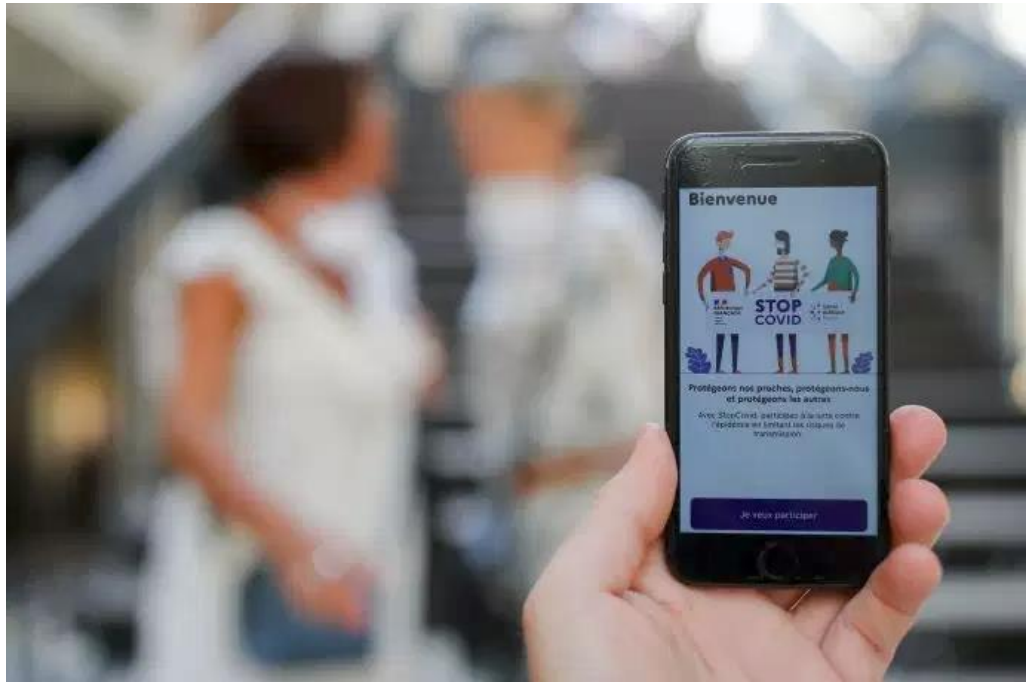
Des coûts mensuels de l'ordre de 100.000 €





Et vous, qu'en pensez-vous ?

- L’avez-vous installée ?
- Allez-vous l’utiliser ? Oui / Non / Pourquoi



Pensez-vous que l’outil a été conçu en suivant une démarche éthique, entre nos droits fondamentaux, l’efficacité sur la lutte contre la pandémie et les impacts potentielles sur les personnes ?

Exemples :

- Surveillance ?
- Vie privée ? Protection des données ?
- Légitimité / efficacité ?
- Accountability ?
- Taux d'utilisation pour être efficace ?
- Incitation à l'usage ?
- Droits fondamentaux ?
- Le Captcha ?

Le pour et le contre

- Participer à endiguer la pandémie
- CNIL consultée et prise en compte de ses recommandations
- Prise de conscience des données et de leur usage
- Protocole de sécurisation « Robert » plutôt robuste mais dévoyé
- 1 vie sauvée, quel qu'en soit le coût
- Installation de l'application sur la base du volontariat

- Collecte + de données qu'annoncé et que nécessaire
- Risque de pratiquer la centralisation des données (TOUS les contacts sont envoyés au serveur)
- Sans un usage massif, utilité réelle de l'application
- Coûts exorbitants : entre 100 et 200 k€ par an pour hébergement et maintenance
- Sans tests massifs, en complément, l'application n'est pas utile.
- L'intégralité du code n'a pas été ouverte.
- Bcp d'autorisations demandées par l'application : localisation géographique, la prise d'une photo

Eclairage éthique

Les interventions traduisent bien le dilemme éthique sur cette application avec une tension entre deux principes (droits fondamentaux) et objectifs :

Protection de la santé vs. protection de la vie privée.

Cela illustre aussi la difficulté d'arbitrer avec des intérêts légitimes et parfois contradictoires. Les questions sur l'efficacité de l'application (et sa démonstration) sont dès lors intéressantes à explorer.

Les principes éthiques (du numérique) de loyauté et de transparence ont aussi été évoqués :

- Loyauté par rapport au traitement des données : sur la collecte de plus d'informations que prévu (au-delà des identifiants des seules personnes contacts de proximité 15 min + 1 m)
- Transparence : sur l'information concernant le fonctionnement de l'application.

Un questionnement a aussi émergé : Pourquoi les bonnes pratiques et les débats sur l'application Stopcovid ne sont pas reproduits pour les applications privées? Enfin pourquoi sommes-nous si sensibles aux bonnes pratiques sur cette application, alors que nous laissons nos réseaux sociaux préférés utiliser commercialement nos données personnelles sans rien y trouver à redire!

Cela peut être un sujet intéressant à explorer, sachant que la loi informatique et libertés s'applique à tous les responsables de traitement de données (privés et publics). Ces derniers ont l'obligation de consulter la CNIL lorsqu'une évaluation d'impacts visée au RGPD met en lumière que des risques élevés peuvent exister par rapport aux traitements des données dans le cadre d'une application.

Merci !