HAL
open science

Données de santé : l'arbre StopCovid qui cache la forêt Health Data Hub

Bernard Fallery

To cite this version:

Bernard Fallery. Données de santé : l'arbre StopCovid qui cache la forêt Health Data Hub. The Conversation France, 2020. hal-03125892

HAL Id: hal-03125892

https://hal.science/hal-03125892v1

Submitted on 9 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Données de santé : l'arbre *StopCovid* qui cache la forêt *Health Data Hub*

Bernard Fallery

Université de Montpellier, MRM

(version complète sur Theconversation.com)

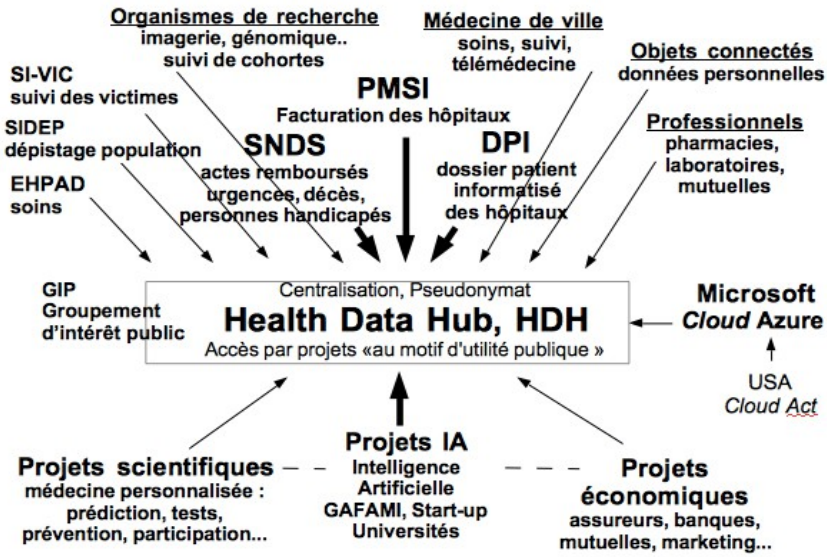
StopCovid, l'arbre qui cache la forêt

Le projet StopCovid pour un traçage social « acceptable » à l'aide des smartphones a focalisé l'intérêt de tous (The Conversation 2020). Apple et Google se léchaient déjà les babines avec un protocole API commun pour l'international. Mais en France une forte controverse devant l'espace public, cumulée avec le retrait de l'Allemagne du protocole API de l'Inria et l'échec constaté à Singapour (20% d'utilisateurs), mèneront à l'abandon de StopCovid. « *Ce n'est pas prêt et ce sera sûrement doucement enterré. À la française* », estimait un député LREM le 27 avril auprès de l'AFP.

Health Data Hub, la foret qui se cache derrière l'arbre

Mais pendant ce temps-là... un projet bien plus large continue à marche forcée, celui de la plateforme des données de santé (en franglais Health Data Hub, HDHub).
Dès la remise du rapport Villani en mars 2018, le Président de la République annonce le projet HDHub. Dès octobre, une mission de préfiguration définit les traits d'un système national centralisé de l'ensemble des données de santé publique, un guichet unique à partir duquel l'intelligence Artificielle IA pourrait optimiser des services de reconnaissance des formes et de prédiction personnalisée (The Conversation 2018).

Mais l'écosystème de l'IA s'apprête aussi à franchir une nouvelle marche : de véritables **données massives** (des hôpitaux, de la recherche, de la médecine de ville, des objets connectés...) et un véritable **marché massif** de la santé (prestigieux et à valeur potentielle énorme avec plus de 12% du PIB). La France avec l'Assurance-maladie et le Royaume-Uni avec le NHS font ici figure de test, puisque des données cohérentes et fiables y sont maintenues depuis des décennies : Amazon a déjà accès à l'API du NHS pour alimenter son assistant vocal, Microsoft a déjà signé l'hébergement de toutes les données de santé françaises (stockage, gestion des logs et des annuaires, puissance de calcul, conservation des clés de chiffrement).



Le projet HDHub mené « au pas de charge »

En novembre 2018, Stéphanie Combes est nommée cheffe de projet. Fin 2018, le choix de Microsoft est déjà acté (en "dispense de marché public"), alors même que les principes du HDHub attendront juillet 2019 (dans la Loi Santé) et que ses missions ne seront définies qu'en avril 2020 (par arrêté ministériel). La CNIL, malgré ses échanges avec S. Combes, continue à se poser de nombreuses questions dans un avis du 23 avril 2020, et notamment : « *La Commission s'interroge donc sur les conditions de démarrage anticipé de la solution technique...* ». D'autres voix s'étaient aussi inquiétées de la gestion si hâtive du projet (comme le Conseil national des barreaux, l'Ordre national des médecins, un député LREM...) et des collectifs avaient lancé des alertes argumentées (comme les professionnels de InterHop, les Entreprises du Logiciel Libre...) ou des vidéos de révolte de certains médecins.

Vidéo Youtube 2:19 d'un médecin révolté, voir le **lien**

Health Data Hub, un cas d'école sur toutes les problématiques du numérique

Contourner l'arbre qui cache la foret, c'est découvrir toute l'étendue des questions posées par la « transformation numérique » dans la société, et ici dans la santé.

Les questions politiques se cristallisent ici autour du choix de Microsoft, que S. Combes justifie très classiquement par l'urgence, sans publication des délibérations : «*Microsoft était le seul capable de répondre à nos demandes... On a préféré aller vite, pour ne pas prendre de retard et pénaliser la France...* » (Mediapart 2020). Une question de politique nationale, déjà soulevée dans The Conversation 2020, puisqu'il s'agit de faire gérer un **bien public** par un acteur privé, et sans espoir de réversibilité. Mais aussi une question politique de **souveraineté** numérique européenne quand cet acteur étasunien se trouve soumis au *Cloud Act*, loi de 2018 qui permet aux juges américains de demander l'accès aux données sur des serveurs situés en dehors des Etats-Unis.

Vidéo Youtube d'un débat avec S. Combes, voir le **lien**

Les questions techniques se révèlent ici dans un vif débat entre centralisation ou inter-opérabilité des bases de données. La **centralisation** définit des architectures de « défense en profondeur » avec des barrières successives (option de défense choisie par exemple dans le nucléaire, dans le projet HDHub cette défense est sous-traitée chez Microsoft). S. Combes observe (TIC-Santé 2019) que « s*i on veut faire du traitement de données à cette échelle, on doit centraliser, c'est la seule solution* ». A l'opposé, la vision technique des architectures de **l'inter-opérabilité** vise à « ne pas mettre ses oeufs dans le même panier » : d'une part la majorité des attaques ne viennent pas de l'extérieur mais de l'intérieur (avec un risque plus élevé en cas de centralisation) et d'autre part l'anonymat ne résiste pas à la ré-identification d'une personne par croisement de données. Cette architecture décentralisée consiste alors à gérer les échanges en réseau entre des bases de données hétérogènes et entre des traitements distribués, par les interfaces aujourd'hui standardisées et ouvertes (une option choisie par exemple dans le projet eHop pour un groupe d'hôpitaux et qui a l'avantage de maintenir localement les compétences des ingénieurs et des soignants, nécessaires à la qualification des données de santé).

Les questions juridiques concernent ici le consentement et le secret médical, à plus long terme la responsabilité individuelle des médecins. Les principes européens du RGPD (The Conversation 2019) organisent le **consentement** dès la conception des systèmes d'information (*privacy by design*) et par une culture de transparence interne dans les organisations (le délégué à la protection des données). Les données de patients touchent bien sûr à leur intimité, mais la durée, le droit de retrait et surtout la finalité claire d'une utilisation de ces données sont des principes intangibles fixés par la CNIL et longuement rappelés dans son avis du 23 avril 2020. Dans sa réponse S. Combes a donné des perspectives : « *En effet, les données ne sont censées être stockées que durant la période de l'état d'urgence sanitaire. À sa fin, elles devront être détruites,* ***sauf si*** *un autre texte prévoit cette conservation lors de la mise en place finale du Health Data Hub»*. Dans la pratique les patients pourraient être soumis à une rupture du **secret médical**, un principe juridique mais aussi une règle éthique qui fonde la confiance basée sur le serment

d'Hippocrate. Une rupture de cette confiance présenterait bien sûr des risques en terme de santé publique.

Les questions économiques se cristallisent autour des enjeux de la transformation numérique. Les tenants d'un néo-libéralisme voient surtout dans le numérique une force de **destruction créative** : une dérégulation et un désengagement des états favorisent l'innovation disruptive (The Conversation 2020) et la croissance par des start-up. Au delà du seul intérêt scientifique, un développement rapide de l'IA grâce aux GAFAMI peut donc être au service « de l'intérêt général », une finalité introduite en 2019 dans la Loi santé. A l'opposé, les tenants d'une alternative économique voient surtout dans le numérique une possibilité de gestion des **communs numériques** (en suivant les analyses de Elinor Ostrom) : des ressources immatérielles non-rivales, dont les règles d'accès et d'usage sont gérées par des acteurs en communautés auto-organisées très diverses (depuis Internet, Wikipédia... jusqu'à l'Open data, le Logiciel libre ou les énormes bases scientifiques du type *Protein Data Bank*... voir le rapport *Sharing Data in a Medical Information Commons*) . On dénonce alors ici l'idée d'une séparation entre la qualification des données médicales (un long travail de collecte et de tri, financé par le secteur public et soumis aux traités de libre circulation des données) et la valorisation de ces données (une marchandisation de la santé, par le secteur privé protégé par les traités sur les brevets).

Au final **la question sociale** du contrôle sanitaire de nos comportements (The Conversation octobre 2019) ne peut pas être analysée sans les concepts forgés par les sociologues. Michel Foucault a décrit le passage progressif à la **Société disciplinaire** avec les concepts de « biopolitique » (de la lèpre à la peste, des formes d'exercice du pouvoir sur les corps, The Conversation 2020) et celui de « gouvernementalité » (associant gouvernement et rationalité, dans des technologies du gouvernement des individus et de soi, pour assurer l'auto-discipline : hier déjà, le confinement, l'école, l'hôpital, les statistiques... maintenant les panoptiques du drone et du bracelet). Gilles Deleuze a alors décrit un nouveau passage vers la **Société de contrôle** par le collier électronique, avec les concepts de « langage numérique » d'accès à la réalité et celui « d'atermoiement illimité » décrit par Kafka (The Conversation décembre 2015) : il ne s'agit plus de discipliner et d'ordonner, mais de contrôler en gérant tout désordre.

Vidéo Youtube d'une présentation de la gouvernementalité par A. Rouvray, voir le **lien**

Aujourd'hui des sociologues comme A. Rouvray ou D. Cassada montrent un prochain passage à la **Société des traces** avec les concepts de gouvernementalité algorithmique (au delà d'une maitrise du probable il s'agit d'une maitrise sur le potentiel lui-même, pour « ajuster » nos comportements) et de sousveillance (non plus une sur-veillance, mais une sous-veillance par un quadrillage discret, immatériel et omniprésent de tous les types de traces que nous laissons : nos signaux, nos productions, nos empreintes, nos passages, nos liens...).