

# COVID-19 : MODÉLISATIONS ET DONNÉES POUR LA GESTION DE CRISES SANITAIRES

RAPPORT DE L'ACADÉMIE DES TECHNOLOGIES

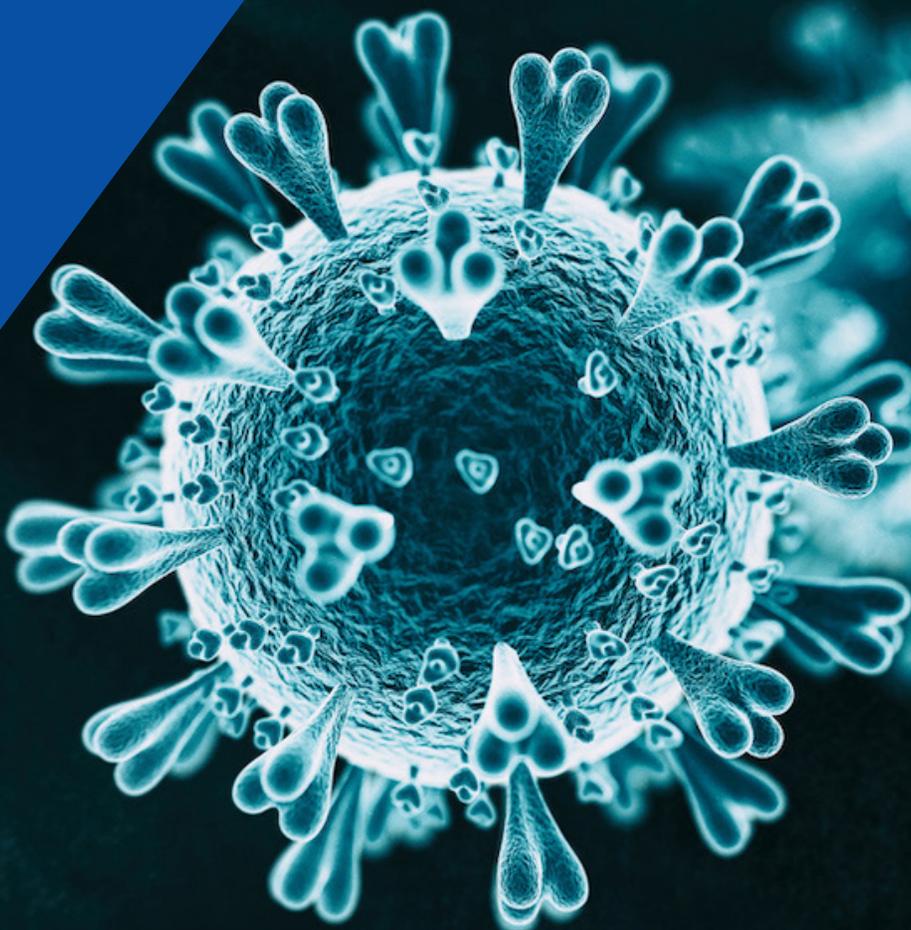


Image de couverture : <https://thefederal.com/file/2020/05/abs-2019ncov-virus-hologram-picture-id1204033162-1.jpg>

ACADÉMIE  
POUR UN PROGRÈS  
RAISONNÉ  
CHOISI  
PARTAGÉ  
DES  
TECHNOLOGIES

COVID - 19 :  
MODÉLISATIONS ET DONNÉES  
POUR LA GESTION DE CRISES SANITAIRES

*Rapport de l'Académie des technologies*

Académie des technologies, pôle numérique

---

**Rédacteurs Albert Benveniste et Laurent Gouzènes**

avril 2021

Académie des technologies  
Le Ponant – Bâtiment A  
19, rue Leblanc  
75015 PARIS  
+33(0)1 53 85 44 44  
[secretariat@academie-technologies.fr](mailto:secretariat@academie-technologies.fr)  
[www.academie-technologies.fr](http://www.academie-technologies.fr)

©Académie des technologies  
ISBN : 979-10-97579-29-6

# SOMMAIRE

<b>OBJET DE CE RAPPORT</b>	<b>3</b>
<b>SYNTHÈSE ET RECOMMANDATIONS</b>	<b>7</b>
Synthèse	7
Recommandations	8
<b>CHAPITRE I — PREMIÈRES LEÇONS DE LA CRISE DE LA COVID-19</b>	<b>15</b>
Covid-19 : une crise globale et inédite	15
Questions relatives à l'organisation	16
La gestion de crise dans le monde industriel ?	21
Quelques observations sur les actions de L'Europe	23
<b>CHAPITRE II — IMPORTANCE DE LA MODÉLISATION</b>	<b>25</b>
Les services pratiques rendus par la modélisation sont précieux	25
Rappel sur les notions de système et de modélisation	26
Domaines de modélisation considérés dans le cadre de la covid-19	28
Pandémies humaines : en quoi cela change la démarche de modélisation	29
<b>CHAPITRE III — DES DONNÉES À LA MODÉLISATION ET À LA MISE EN ŒUVRE NUMÉRIQUE</b>	<b>35</b>
La question des données	35
Modélisation : domaines et outils mathématiques	42
Mise en œuvre numérique	48
Les forces industrielles en Europe	51

<b>CHAPITRE IV — LEÇONS DES ENTRETIENS RÉALISÉS :</b>	
<b>CE QU'IL EST POSSIBLE DE FAIRE</b>	<b>53</b>
Modélisation et données en sont la colonne vertébrale	<b>53</b>
Plates-formes, approche cockpit, jumeau numérique, données : état de l'art	<b>55</b>
Capitaliser l'expérience et les meilleures pratiques	<b>57</b>
<b>CONCLUSION — POUR UNE PLATE-FORME DE MODÉLISATION POUR LES CRISES SANITAIRES</b>	<b>59</b>
Combattre l'image de « décisions-girouette »	<b>60</b>
Prendre en compte l'ensemble des éléments pour faire émerger la meilleure stratégie	<b>61</b>
Donner les éléments explicatifs de la décision	<b>61</b>
Avec quelle organisation ?	<b>61</b>
S'appuyer sur les acquis industriels et sur la recherche	<b>62</b>
<b>ANNEXES</b>	<b>63</b>
Annexe 1 — Entretien avec des membres de la société Thales	<b>64</b>
Annexe 2 — Entretien avec Dassault Systèmes	<b>82</b>
Annexe 3 — Entretien avec Pierre Haren, CausalityLink	<b>92</b>
Annexe 4 — Entretien avec Alex Fleischer et Alain Chabrier IBM	<b>108</b>
Annexe 5 — Entretien avec Jean-Michel Lasry — <b>Mean Field Games</b> (MFG)	<b>118</b>
Annexe 6 — Entretien avec Patrick Guérin, Openhealth :	
Unité COMETE : Analyse des eaux usées	<b>128</b>
Annexe 7 — Entretien avec Patrick Lagadec	<b>136</b>
Annexe technique et bibliographique :	<b>140</b>
1. État de la pandémie	<b>140</b>
2. Sur la modélisation épidémiologique	<b>148</b>
3. La question des données	<b>154</b>
4. Aspects économiques, sociaux et sociétaux	<b>162</b>

## OBJET DE CE RAPPORT

**L**a crise de la covid-19 a rapidement dépassé le domaine strictement médical pour impacter l'ensemble de l'économie et des activités françaises. Le système « France » a été soumis à une série de chocs majeurs successifs (santé, confinements, économie, psychologie...), qui ont mis en évidence les multiples interactions entre des composantes qui fonctionnaient habituellement dans un équilibre invisible et qui a été gravement perturbé. Ceci vaut également pour les autres pays européens.

Relever le défi de la santé dans un environnement fortement perturbé s'est imposé au gouvernement et aussi à tous les acteurs de l'économie : comprendre le présent, anticiper le futur, comprendre les leviers d'actions possibles, étudier des scénarii sont devenus des éléments clés pour chacun à son échelle.

La modélisation et la simulation ont ainsi acquis une nouvelle visibilité liée à ces besoins. Des limitations sont cependant apparues dans la seule prise en compte de la dynamique de l'épidémie. D'autres facteurs se sont révélés essentiels (industriels, économiques, sociétaux, psychologiques).

Il faut jeter les bases d'une véritable politique de gestion de crise en France et en Europe, fondée sur les données et les modèles. La France et l'Europe sont richement dotées en ressources scientifiques et industrielles pour mener une telle politique. Les approches pluridisciplinaires en matière de modélisation sont indispensables et à encourager, y compris autour de l'épidémiologie, et nos forces en mathématique et informatique, tant académiques qu'industrielles, doivent être mobilisées dans ce but.

**Avertissement** : les aspects médicaux de la maladie : virus, traitements, vaccins, etc., sont en dehors du champ de ce rapport. Ils sont amplement étudiés par les communautés scientifiques qui se rattachent au secteur de la santé.

## COMMENT LIRE CE RAPPORT ?

- Les premières pages (9 pages) contiennent les recommandations du rapport ;
- Les chapitres I à V (40 pages) contiennent les analyses réalisées :
  - chapitre I : La problématique générale de la gestion de crise et premiers retours d'expérience sur la crise 2020 ;
  - chapitre II : L'importance de la modélisation pour la gestion de crise : réaliser des prévisions rapidement pour optimiser les décisions de terrain et éviter les erreurs ;
  - chapitre III : Les problématiques scientifiques et informatiques à considérer : collecte des données, recherche opérationnelle, théorie des jeux, fouille de textes et causalités, plateformes informatiques ;
  - chapitre IV : les enseignements principaux issus des entretiens et de la documentation scientifique et technique et les besoins à satisfaire pour disposer d'outils performants pour la gestion de crise : des outils souples et évolutifs permettant dynamiquement d'agréger les données et d'évaluer des scénarii dans des dimensions très variées.

*Nous insistons sur l'importance des annexes (p 59 à la fin) : elles constituent les sources d'information principales pour l'écriture de ce rapport.*

### « le « ou « la « covid ?

L'académie française a publié le long explicatif suivant :

« Covid est l'acronyme de corona virus disease, et les sigles et acronymes ont le genre du nom qui constitue le noyau du syntagme dont ils sont une abréviation. On dit ainsi la S.N.C.F. (Société nationale des chemins de fer français) parce que le noyau de ce groupe, société, est un nom féminin, mais le C.I.O. (Comité international olympique), parce que le noyau, comité, est un nom masculin. Quand ce syntagme est composé de mots étrangers, le même principe s'applique. On distingue ainsi le FBI, Federal Bureau of Investigation, « Bureau fédéral d'enquête », de la CIA, Central Intelligence Agency, « Agence centrale de renseignement », puisque dans un cas on traduit le mot noyau par un nom masculin, bureau, et dans l'autre, par un nom féminin, agence. Corona virus disease – notons que l'on aurait pu préférer au nom anglais disease le nom latin morbus, de même sens et plus universel – signifie « maladie provoquée par le corona virus ("virus en forme de couronne") ». On devrait donc dire la covid 19, puisque le noyau est un équivalent du nom français féminin maladie. Pourquoi alors l'emploi si fréquent du masculin le covid 19 ? Parce que, avant que cet acronyme ne se répande, on a surtout parlé du corona virus, groupe qui doit son genre, en raison des principes exposés plus haut, au nom masculin virus. Ensuite, par métonymie, on a donné à la maladie le genre de l'agent pathogène qui la provoque. Il n'en reste pas moins que l'emploi du féminin serait préférable et qu'il n'est peut-être pas trop tard pour redonner à cet acronyme le genre qui devrait être le sien. »

Parlant du virus nous adoptons donc le masculin : le covid, et, parlant de la maladie le féminin : la covid 19,



## SYNTHÈSE ET RECOMMANDATIONS

### SYNTHÈSE

La crise due à la covid-19 a rapidement dépassé le domaine strictement médical pour impacter l'ensemble de l'économie et des activités françaises. Comme partout dans le monde, le système « France » a été soumis à une série de chocs majeurs successifs (santé, confinements, économie, psychologie...), qui ont mis en évidence les multiples interactions entre des composantes qui fonctionnaient habituellement dans un équilibre invisible qui a été gravement perturbé.

Relever le défi de la santé dans un environnement fortement perturbé s'est imposé au gouvernement et aussi à tous les acteurs de l'économie : comprendre le présent, anticiper le futur, comprendre les leviers d'actions possibles, étudier des scénarii sont devenus des éléments clés pour chacun à son échelle.

La modélisation et la simulation ont ainsi acquis une nouvelle visibilité liée à ces besoins. Des limitations sont cependant apparues dans la seule prise en compte de la dynamique de l'épidémie. D'autres facteurs se sont révélés essentiels (industriels, économiques, sociétaux, psychologiques), et ce parce que la pandémie était humaine, et non pas animale ou végétale.

La présente étude n'aborde pas les aspects biologiques (structure du virus, mécanismes d'actions, qualité des vaccins, etc.). Elle se concentre sur les autres points, liés à la collecte de données, à la modélisation, à l'interaction entre éléments physiologiques et physiques, les réactions de populations humaines et, plus généralement, à tous les domaines qui relèvent des mathématiques, de l'informatique et de la théorie de la décision.

---

---

**Pour gérer une crise de grande ampleur telle que la COVID, il faut, à la fois, avoir anticipé, et être armé pour improviser en obtenant néanmoins de bons résultats : une approche système appuyée sur données et modèles est le bon investissement pour l'anticipation, et l'arme de choix pour une improvisation efficace. Il faut donc jeter les bases d'une véritable politique de gestion de crise en France et en Europe, fondée sur une approche système appuyée sur données et modèles.**

**La France et l'Europe sont richement dotées en ressources scientifiques et industrielles pour une telle politique.**

**Les approches pluridisciplinaires en matière de modélisation sont indispensables et à renforcer, et nos forces en mathématique et informatique, tant académiques qu'industrielles, doivent être mobilisées dans ce but.**

---

---

## RECOMMANDATIONS

### **D**isposer de plates-formes pour l'intégration de modélisations d'origines et de natures différentes

Un point clé dans la gestion de crises inédites est que l'on découvre au cours de la crise quels sont les facteurs essentiels qui interviennent et sur lesquels les autorités peuvent interagir. Il est donc illusoire d'espérer échapper à l'improvisation. Mais *improviser ne s'improvise pas* : il faut être pleinement préparé et outillé.

---

---

**Il faut donc disposer d'une plate-forme permettant l'intégration rapide et à coût réduit de modélisations d'origines et de natures différentes et des données associées — en quelque sorte un "Crisis Model Hub". Cela peut être envisagé, soit sur une base nationale, soit au niveau européen.**

---

---

(Un peu à la manière dont sont structurées aujourd'hui les plates-formes de *big data*.)

De telles plates-formes existent en Europe avec des offres de premier plan mondial, mais restent peu connues hors des secteurs industriels pour lesquels elles ont été initialement conçues.

---



---

**Là déclinaison de ce type de plateforme pour la gestion de crises sanitaires suppose l'accès à des données publiques, une restitution publique des résultats fournis par ces plateformes, ainsi qu'une certaine forme de publication des modèles invoqués, prolongeant et approfondissant vers d'autres domaines le concept de Health Data Hub existant pour le domaine de la santé. Pour assurer la souveraineté et l'indépendance de l'exploitation des données et des prédictions proposées par les modèles, l'initiative Franco-Allemande GAIA-X constitue un cadre naturel à vocation européenne.**

---



---

La question suivante reste ouverte à ce stade : dans le cadre de l'utilisation pour une crise sanitaire de type covid, par qui devrait être pilotée cette plate-forme <sup>1</sup>?

- Commission européenne
- services du Premier ministre (comme Secrétariat général de la défense et de la sécurité nationale ?
- sécurité civile ?
- cellule interministérielle de crise ?
- une délégation *ad hoc* de service public ?

Et, sur un plan plus général, concernant la coordination européenne, on pourrait envisager une action conjointe, missionnée par les pouvoirs publics, de l'acatech allemande et de l'Académie des technologies française sur ce dossier (les deux pays étant fortement positionnés sur le plan industriel sur ce sujet).

## **P**ermettre l'accès aux données d'intérêt général

---



---

**Là comme pour bien d'autres situations, il faut garantir et organiser l'accès aux données d'intérêt général.**

---



---

On parle ici des données d'intérêt général au sens du rapport Bothorel<sup>2</sup>. Dans le cas de la crise due à la covid-19 les données concernées incluent : santé publique, localisation, transports, etc. Cette mise à disposition comporte des aspects légaux (protection de la vie privée), de coût (pour l'accès), de délais et d'actualité des données.

Notons l'annonce par le gouvernement, en mars 2021<sup>3</sup>, de l'ouverture de soixante jeux de

---

1 Informations détaillées sur GAIA-X, voir : <https://www.data-infrastructure.eu/GAIA-X/Navigation/EN/Home/home.html> Pour le Health Data Hub, voir <https://www.health-data-hub.fr/>

2 Voir <https://www.tresor.economie.gouv.fr/Articles/2020/12/28/rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources-le-reseau-international-du-tresor-en-premiere-ligne>

3 <https://www.vie-publique.fr/discours/278923-amelie-de-montchalin-05032021-numerisation-des-services-publics>

données et codes sources [...] dont notamment le code source de FranceConnect<sup>4</sup> [...]. Cette orientation va dans le bon sens elle gagnera sans doute à être précisée, et le présent rapport propose quelques pistes.

## **C**onduire et archiver les retours d'expérience

---

---

**Il convient de mener, après la crise, une étude (benchmark) sur comment d'autres pays se sont organisés pour gérer l'épidémie.**

**Il convient d'archiver les expériences de crises achevées (par exemple en s'appuyant sur les plates-formes intégratives de modélisation).**

---

---

## **F**avoriser les recherches transdisciplinaires en modélisations mathématiques et informatiques

L'épidémiologie était un domaine établi depuis longtemps, mais connu des seuls spécialistes, jusqu'à ce que la crise de la covid 19 la jette sous les lumières de l'actualité. Un enjeu central de la modélisation des épidémies humaines (surtout si la modélisation doit servir à prendre des décisions de politique sanitaire), c'est d'incorporer à l'intérieur du modèle le contexte socio-économique et réglementaire.

---

---

**Il faut favoriser l'émergence d'une « nouvelle » école de modélisation d'épidémiologie humaine, qui produira des modèles mélangeant les dimensions de comportements sociaux culturels, les anticipations, et l'épidémiologie classique dont l'objet est la propagation du virus.**

---

---

4 <https://franceconnect.gouv.fr/>

## **P**révoir des formations

La pluridisciplinarité s'impose dans les sujets complexes.<sup>5,6</sup>

---

---

**Il faut développer la formation dans les domaines suivants :**

- gestion de crise
- théories d'aide à la décision
- modélisation et systèmes complexes.

**Les niveaux de formation peuvent aller de la sensibilisation à l'expertise.**

---

---

---

---

**Nous recommandons la mise sur pied de formations permettant à des non-experts en modélisation de faire un usage correct des résultats de ces modèles.**

---

---

Dans la mesure où les résultats fournis par les modèles sont utilisés par des non-experts des dits modèles<sup>7</sup>, il est important que soit délivrée une formation pour non-spécialiste, destinée à éviter les interprétations et usages erronés de ces résultats.

5 <https://formation.cnam.fr/rechercher-par-discipline/gestion-des-crisis-et-situations-urgentes-et-exceptionnelles-en-securite-sanitaire-208430.kjsp>

6 Par exemple à l'École polytechnique.

7 S'agissant de la covid-19 et de l'épidémiologie, on peut compter parmi ces non-experts la plupart des décideurs politiques et une bonne partie du corps médical.

## LE GROUPE DE TRAVAIL AYANT PRODUIT CE RAPPORT

Le groupe de travail a commencé ses travaux en juin 2020 et terminé une première version du rapport en février 2021.

### Rédacteurs :

ALBERT BENVENISTE, chercheur émérite à Inria, (*Académie des technologies*)

LAURENT GOUZÈNES, KM2 Conseil (méthodes et outils de gestion de la connaissance, génération automatique de documents).

### Ont contribué à la rédaction

ALAIN BENSOUSSAN (*Académie des technologies*)

GÉRARD ROUCAIROL (*Académie des technologies*)

MAKREM DRIDI

MARKO ERMAN (*Académie des technologies*)

PIERRE HAREN (*Académie des technologies*)

PATRICK JOHNSON (*Académie des technologies*)

BERNARD LARROUTUROU.

JEAN-MICHEL LASRY,

### Membres du groupe de travail

STÉPHANE ANDRIEUX (*Académie des technologies*)

YVES BAMBERGER (*Académie des technologies*)

ALBERT BENVENISTE (*Académie des technologies*)

PIERRE-ÉTIENNE BOST (*Académie des technologies*)

ANGE CARUSO,

YVES CASEAU (*Académie des technologies*)

MARKO ERMAN, (*Académie des technologies*)

LAURENT GOUZÈNES

PIERRE HAREN, (*Académie des technologies*)

GÉRARD ROUCAIROL (*Académie des technologies*)

BRUNO STOUFFLET (*Académie des technologies*)

BERNARD TARDIEU (*Académie des technologies*)

DOMINIQUE VERNAY (*Académie des technologies*)

### Entretiens conduits auprès de

*Dassault-Systèmes*

MAHEL ABAAB-FOURNIAL

PATRICK JOHNSON (*Académie des technologies*)

VINCENT SIESS

## Thales

MAKREM DRIDI

MARKO ERMAN (*Académie des technologies*)

LAURENT LETELLIER

HERVE NOEL

## Causality Link :

PIERRE HAREN (*Académie des technologies*)

ERIC JENSEN

CHRISTIAN DEUTSCH

## IBM

ALAIN CHABRIER

ALEX FLEISCHER

## OpenHealth

PATRICK GUÉRIN

## Autres

ALAIN BENSOUSSAN, mathématicien, professeur émérite à l'université de Paris Dauphine et professeur à l'université du Texas à Dallas (*Académie des technologies*)

PATRICK LAGADEC, consultant indépendant (spécialiste de la gestion de crises)

BERNARD LARROUTUROU - CGEDD/TECI (gouvernement, développement durable)

JEAN-MICHEL LASRY, mathématicien, professeur émérite à l'université de Paris Dauphine, et Institut Louis Bachelier

Le groupe de travail a discuté des objectifs, puis participé aux premières relectures. Les contributeurs ont apporté du contenu à ce rapport par des séances de travail en sous-groupes. Les entretiens ont été conduits avec échanges approfondis et démonstrations et des contributions au-delà des entretiens en ont résulté. Tout le travail s'est déroulé en téléconférence.

Une relecture détaillée a été effectuée par Christian Saguez (*Académie des technologies*).

### Éléments concernant les déclarations d'intérêt

- Les experts industriels interviewés l'ont été au titre de représentants de leur employeur.
- Pour les deux rédacteurs et contributeurs hors de la liste des experts industriels interviewés, il n'est déclaré aucun lien d'intérêt avec les compagnies ayant fait objet d'entretiens.



## Chapitre I

# PREMIÈRES LEÇONS DE LA CRISE DE LA COVID-19

### COVID-19 : UNE CRISE GLOBALE ET INÉDITE

La crise due à la covid-19 est passée en quelques semaines d'une maladie à traiter individuellement, à une pandémie qui a mis à mal les systèmes de santé nationaux, puis des économies entières dans une chute de dominos qui a démontré une nouvelle fois l'interdépendance liée à la mondialisation :

- **une crise globale, affectant toutes les régions du monde** (point n'est besoin de commenter) ;
- **une crise globale, affectant la société sous toutes ses facettes** : Le fait que la crise due à la covid-19 affecte tous les aspects de nos sociétés est valable dans toutes les régions du monde, avec des impacts même là où les difficultés sanitaires ont été faibles ou modérées. L'impact économique et social est, en grande partie, dû aux changements de comportement des populations dans les pays les plus riches, avec un fort effet sur le tourisme<sup>8</sup> ;
- **une crise profondément inédite... de mémoire d'homme** : l'Occident avait oublié les grandes pestes du Moyen-Âge et la grippe espagnole de 1918-1920 et ne sentait que peu concerné par les pandémies récentes touchant le tiers-monde (dont Ebola) ;
- **le monde n'y était pas préparé** : La Chine a mis du temps à admettre la gravité de la pandémie (les raisons peuvent en être en partie politiques, mais pas totalement : la surprise semble avoir été réelle). Le reste du monde, et tout particulièrement le monde occidental, a mis du temps à reconnaître que l'épidémie se disséminait. On a constaté une impréparation logistique variable selon les pays. En France, elle a concerné principalement masques, respirateurs, et tests ;

---

8 Pour ce qui concerne la France, le site <https://www.insee.fr/fr/information/4479280> de l'Insee tient à jour les effets de l'épidémie sur la société Française. Un portrait social de la France est publié annuellement : <https://www.insee.fr/fr/statistiques/fichier/4928952/FPS2020.pdf> pour décembre 2020.

- **la difficile appréciation des services fournis par l'État** en cette période de pandémie. La pandémie affecte aussi profondément la capacité à en mesurer les effets : les instruments pour cela n'ont pas été prévus pour une période de crise<sup>9</sup> ;
- **une question inédite sur les leviers de pilotage** : les leviers de pilotage à disposition des pouvoirs publics pour le contrôle de la pandémie ont été et demeurent les suivants :
  - promotion des règles sanitaires (gestes barrière...) contraintes de type confinement (fermeture de telle ou telle activité, télétravail, couvre-feu...),
  - outils de santé publique : tests, identification et traçage des personnes infectées, vaccination.

La première catégorie de leviers est de type réglementaire : ce sont les leviers de cette catégorie qui ont principalement un effet négatif sur l'économie et la société.

Les deux catégories de leviers sont sensibles à la question de l'acceptabilité par la société. De ce point de vue, des différences importantes sont apparues entre, d'une part l'Asie du Sud-Est et l'Extrême-Orient en général, et, d'autre part, le monde occidental. On rencontre à la fois des impatiences concernant les tests, et aujourd'hui les vaccins, et des réticences à leur sujet. Rappelons que la campagne de test massif de toute la population mis en place dans la ville du Havre<sup>10</sup> a eu un succès effectif mitigé, avec 11 % de la population ayant répondu.

Au total, une crise de type covid-19 était à la fois prévisible (citée dans tous les rapports sur le H1N1 — voir notamment<sup>11</sup>) — et imprévue (quand, comment ?), mais la préparation et la mise en œuvre en France ont été très insuffisantes (cf. rapport Sénat<sup>12</sup>). La partie imprévisible (dans son développement holistique) découle en grande partie de l'impréparation de l'amont.

## QUESTIONS RELATIVES À L'ORGANISATION

Les questions d'organisation et de préparation à la gestion de l'épidémie sont majeures.

Pour rappeler une expérience récente restée dans toutes les mémoires, l'accident de la centrale de Fukushima<sup>13</sup> révéla que l'opérateur TEPCO, s'il était parfaitement focalisé sur la qualité du nucléaire et la maintenance des centrales, avait complètement omis de prendre en compte les risques externes, comme celui du tsunami<sup>14</sup>.

9 <https://blog.insee.fr/sante-education-services-administratifs-la-difficile-mesure-des-activites-non-marchandes-en-temps-de-crise-sanitaire/>

10 Voir : <https://www.lehavre.fr/actualites/covid-19-testez-vous-pres-de-chez-vous> mais aussi <https://www.lefigaro.fr/sciences/covid-19-au-havre-le-depistage-massif-n-aura-pas-de-suite-20201223>

11 [https://www.gouvernement.fr/sites/default/files/risques/pdf/plan\\_pandemie\\_grippale\\_2011.pdf](https://www.gouvernement.fr/sites/default/files/risques/pdf/plan_pandemie_grippale_2011.pdf)

12 [https://www.senat.fr/commission/enquete/gestion\\_de\\_la\\_crise\\_sanitaire.html](https://www.senat.fr/commission/enquete/gestion_de_la_crise_sanitaire.html)

13 Voir [https://fr.wikipedia.org/wiki/Accident\\_nucl%C3%A9aire\\_de\\_Fukushima](https://fr.wikipedia.org/wiki/Accident_nucl%C3%A9aire_de_Fukushima)

14 De fait, la commission d'enquête du parlement Japonais conclut : Bien que déclenché par ces événements cataclysmiques, l'accident qui s'est ensuivi à la centrale nucléaire de Fukushima Daiichi ne peut pas être regardé comme un désastre naturel. Ce fut un désastre profondément causé par l'homme – qui aurait pu et aurait dû être prévu et prévenu. Et ses effets auraient pu être atténués par une réponse humaine plus efficace.

À l'inverse, et pour considérer un exemple positif directement lié au covid, la gestion de la crise de la covid-19 à Taiwan<sup>15</sup> a été conduite par la cellule *Central Epidemic Command Center* du ministère de la santé, avec un succès reconnu.

Un fait entre tous s'impose lorsque l'on écoute les analyses et recommandations des spécialistes en gestion de crise. Si la crise est prévisible avec des modalités connues à l'avance (c'est la majeure partie des cas de crise en milieu industriel), des plans précis peuvent être prévus et des marges de sécurité peuvent être incluses dans le système pour assurer sa résilience. En revanche, pour les crises dont la nature et les caractéristiques n'ont pas été prévues et analysées au préalable, il est impossible d'échapper à une certaine dose d'improvisation. Mais :

***L'improvisation ne s'improvise pas, sauf à engendrer le chaos.***

Il convient donc d'être préparé pour bien improviser. Ceci concerne les aspects de l'organisation, mais aussi la question de l'outillage pour décupler l'efficacité.

Une hypothèse de base pour la récupération d'un système S en crise est qu'il est plongé dans un sur-système C (grand vis-à-vis de S) qui contient des ressources pour créer la résilience et rétablir le système. La Chine a réussi son confinement car S = « la région de Wuhan » est une petite partie de C = « la Chine », et donc des ressources extérieures ont pu être mobilisées. La difficulté de la crise COVID française (et d'autres pays) est que l'on est parti de S = « le système de santé pris individuellement » et C = « la France » pour arriver très rapidement à S = C = « la France », et donc une situation inédite et majeure.

**Nous rapportons dans cette section, avec prudence, les éléments que nous avons pu collecter concernant l'état de l'administration et du gouvernement français en matière organisationnelle, telle qu'impliquée dans la gestion de la crise de la covid-19.**

### ***La France, son gouvernement et ses administrations malade des « silos » ?***

Il ressort de nos entretiens avec quelques industriels (voir les annexes) que la perception de l'organisation des pouvoirs publics lors de la crise de la covid-19 a été celle d'un travail « en silos ». Cela s'est, en particulier, manifesté par le fait qu'au début de la crise, les consultations auprès d'acteurs industriels ont été entreprises indépendamment par les divers ministères. La coordination entre ministères est perçue comme s'étant déroulée selon une structure pyramidale, avec une consultation du Conseil scientifique et une centralisation au niveau du Conseil de défense et sécurité (organe dont le rôle a été renforcé pour l'occasion).

Avec toute la prudence requise lorsqu'on avance ce type de jugement, plusieurs consultations conduites lors de la rédaction de ce rapport suggèrent que, en effet, en France, l'organisation très pyramidale et le fonctionnement en silos sont courants dans les ministères et les adminis-

15 Voir [https://fr.wikipedia.org/wiki/Pand%C3%A9mie\\_de\\_covid-19\\_%C3%A0\\_Ta%C3%AFwan](https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_covid-19_%C3%A0_Ta%C3%AFwan)

trations, à des degrés variables toutefois (avec un mieux pour le ministère de l'économie, des finances et de la relance).

### **Un organe transversal : la cellule interministérielle de crise**

*(Cette sous-section reprend des éléments qui résultent d'un entretien conduit avec un membre de cette cellule pendant une part significative de la crise<sup>16</sup>.)*

Cette cellule a probablement été activée dans le courant de février 2020. En fait, cette cellule est en veille permanente<sup>17</sup>. Comme toujours, ces cellules n'ont pas de pouvoir décisionnel sur le plan politique, mais ont seulement un rôle d'analyse et de proposition. Par exemple, cette cellule a pu alimenter le Conseil de défense ou des réunions interministérielles avec pour résultat des décisions, ou bien de nouvelles questions. Pour la gestion de la covid-19, trois composantes ont été impliquées :

- le ministère de la santé pour les aspects sanitaires ;
- le ministère de l'Intérieur, qui est traditionnellement en charge de l'interministériel en cas de crise, et qui, donc, coordonne ;
- le Secrétariat général de la défense et de la sécurité nationale, qui est, par essence, qualifié pour la gestion de certaines crises et qui a dans son périmètre d'attention la protection des infrastructures et entreprises essentielles de la nation.

Au début, ces trois composantes étaient séparées et communiquaient par officiers de liaison. Par la suite, les deux premières ont été regroupées dans une cellule unique, avec un commandement unique, mais sans regroupement physique. Pour la composante « Intérieur », une petite centaine de personnes ont été impliquées, avec des activités classiques de gestion de crise : la veille, la logistique, la préparation de scénarios ou la préparation de décisions, notamment à destination des instances politiques.

Vu de l'extérieur (en particulier par des industriels sollicités par divers organes gouvernementaux), un sentiment de fonctionnement « en silo » a cependant émergé. Ce sentiment ne semble pas corroboré par les acteurs de l'intérieur : l'autonomie de chaque ministère dans son champ d'intervention est la tradition et semble très raisonnable, une crise telle que celle de la covid-19 a demandé du transversal, qui a justement été conduit par la cellule interministérielle de crise, le Conseil scientifique, le cabinet du Premier ministre et in fine le Conseil de défense.

### **Des initiatives intéressantes et la difficulté de les multiplier**

Des initiatives intéressantes, initiées de manière plutôt décentralisée, ont vu le jour.

On peut citer en exemple les deux projets d'analyse des eaux usées, les réseaux COMETE<sup>18</sup> et Obépine<sup>19</sup>.

L'unité COMETE, créée par la brigade des marins-pompiers de Marseille, a pour objectif la détection précoce et le pistage sur le terrain du covid, dans une optique de prévention et

16 La personne interviewée a souhaité garder l'anonymat.

17 D'autant plus que le pays sortait tout juste de la crise des gilets jaunes.

18 Cf. annexe sur le réseau COMETE et entretien avec P Guérin.

19 <https://www.reseau-obepine.fr>

soins. Un réseau fédère des acteurs publics (sapeurs-pompiers, des municipalités et des régions) autour d'une plate-forme de collecte de données et de visualisation, avec une très forte réactivité. Voir l'annexe 6.

Pour la recherche, on peut noter l'appui accordé, par une initiative du Mesri (Enseignement supérieur et recherche), au réseau [Obépine](#) regroupant des chercheurs biologistes, mathématiciens, physiciens et spécialistes des capteurs pour mettre sur pied un instrument original du suivi de la pandémie à travers l'analyse des eaux usées sur l'ensemble du territoire<sup>20</sup>. On a vu, depuis, que cette méthode de mesure était particulièrement intéressante par son pouvoir d'anticipation. La mise sur pied de cette action originale s'est faite, non sans effort, conjointement entre le MESRI et les ministères de l'environnement et de la santé — en outre, il y a eu mobilisation d'acteurs publics et privés de la santé (ARS), de la gestion de l'eau et de certaines collectivités locales.

Ces deux réseaux sont un bel exemple de succès de sortie des silos.

Nous devons aussi mentionner l'initiative CoData<sup>21</sup>. Cette initiative a été lancée dès le début de l'épidémie par des entreprises, informaticiens et *data scientists*, qui ont proposé leur support technique pour des travaux de collecte et analyse de données, soit environ 140 personnes et une soixantaine d'entreprises. Il a été proposé une quinzaine de projets opérationnels. Les projets proposés ont été de nature soit à aider le dispositif de santé (collecte de données, aide à la modélisation...), soit à développer des outils plus liés à gestion de la crise hors champ médical (éducation, support à la gestion des aides...). Presque un an plus tard, il semble que la dynamique initiale se soit essoufflée, notamment du côté du privé, pour des raisons financières évidentes, mais peut-être aussi par insuffisance de mécanismes réactifs pour générer des projets utiles et viables de la part des collectivités. Cependant l'organisation GENCI, qui supporte les supercalculateurs nationaux, permet actuellement, par une allocation de ressources prioritaires, le développement de nombreux (38) projets de recherche covid-19 utilisant du calcul intensif (simulation de biologie, screening de molécules, etc.). Il semble, malheureusement, que le déroulement suivi pour les masques (mobilisation d'acteurs spécialistes, fabrication autonome, diffusion, absence de *business model*, abandon...) se soit aussi produit dans ce cadre.

Enfin, on a pu relever que les nouveaux outils pour la Défense prennent en compte le partage d'informations en premier lieu, ceci afin d'accélérer les communications entre équipes et améliorer la qualité des réponses et actions.

### **La société française et sa culture de société**

La politique de communication a sans aucun doute été un point délicat, et elle a rencontré des difficultés. Sur les masques et les tests, certaines communications se sont, *a posteriori*,

20 Voir la page de présentation <https://www.reseau-obepine.fr/presentation-du-reseau-obepine/>. L'Académie des technologies s'est exprimée publiquement dès avril 2020 pour recommander la surveillance des eaux usées à des fins de détection du COVID, se fondant sur les travaux disponibles pour soutenir ce qui est devenu le programme Obépine : <https://www.academie-technologies.fr/blog/categories/publications-de-l-academie/posts/presence-et-activite-du-coronavirus-sars-cov-2-dans-les-eaux-usees>.

21 Voir <https://www.collectif-codata.fr>

révélées malheureuses. Cela dit, le facteur principal, et de plus en plus avec le prolongement de la crise, est l'acceptation par la population des mesures de contrôle sanitaires. Depuis l'automne 2020, c'est devenu l'élément dominant d'incertitude pour le gouvernement, dans l'élaboration des décisions : ainsi, les épidémiologistes expliquent très bien que les contraintes sont d'autant plus efficaces qu'elles sont imposées tôt (dans la partie faiblement croissante du développement de l'épidémie), tandis que le critère d'acceptabilité conduit à différer ces mesures jusqu'au moment où le gouvernement a la conviction que la population les considère comme inéluctables.

*Ce compromis se gère à l'estime, car les autorités ne disposent aujourd'hui d'aucun outil de modélisation permettant vraiment de lier ces deux facteurs et donc de les pondérer.*

### **La gestion de crise en objet d'étude scientifique**

Il existe un champ d'études scientifiques et sociologiques important consacré à la gestion de risques<sup>22,23</sup>, intéressant directement des secteurs régaliens (diplomatie, défense, intérieur) et industriels (logistique, sécurité, etc.). La France dispose ainsi de plusieurs centres d'études avec des professionnels expérimentés. L'idée n'est pas de résoudre un problème donné, mais d'apprendre aux responsables les réflexes à développer pour maîtriser une crise :

- passer l'étape psychologique initiale de stupéfaction ;
- apprendre à s'organiser, notamment dans des contextes désorganisés ou exigeant des canaux nouveaux de communication ;
- apprendre à recueillir les informations, identifier les leviers d'action, etc.

L'importance de la capitalisation et du retour d'expérience est constamment soulignée<sup>24</sup>.

### **Impression d'ensemble**

L'impression générale ressentie par la population française est celle d'une certaine inefficacité lors d'opérations nécessitant la mobilisation de l'administration : pour les masques, les tests et, aujourd'hui, les vaccins. L'anticipation est faible et le démarrage de la réponse est lent (même si les choses ont finalement été mises en place). Les « sous-projets » de la gestion de crise (les masques, les vaccins) n'ont pas donné l'impression d'être traités comme des projets confiés à une organisation spécifique (qui aurait pu être l'armée), mais d'être traités par les multiples canaux habituels du fonctionnement de notre pays, qui ne sont sans doute pas outillés pour et fonctionnent par trop en « silos ».

22 Par exemple [https://portail.polytechnique.edu/i3\\_crg/fr/le-systeme-hospitalier-francais-dans-la-crise-covid-19-une-contribution-des-sciences-de-gestion](https://portail.polytechnique.edu/i3_crg/fr/le-systeme-hospitalier-francais-dans-la-crise-covid-19-une-contribution-des-sciences-de-gestion)

23 Cf IHEDN <https://www.ihedn.fr/type-de-formation/gestion-des-crisis-international> - [https://www.formation-communication-de-crise.fr/formation-communication-de-crise-gestion/?gclid=Cj0KCQjw9\\_mDBhCGARIsAN3PaFPW27F0PtCDzKqRIInksJvnyZk\\_CTSS1u6HiqH6bNODwvz2t8YvXDooaAux7EALw\\_wcB](https://www.formation-communication-de-crise.fr/formation-communication-de-crise-gestion/?gclid=Cj0KCQjw9_mDBhCGARIsAN3PaFPW27F0PtCDzKqRIInksJvnyZk_CTSS1u6HiqH6bNODwvz2t8YvXDooaAux7EALw_wcB)

24 Cf site de Patrick Lagadec <http://www.patricklagadec.net/fr/>

Le groupe ayant rédigé ce rapport avoue partager ces impressions. Lors de la collecte des éléments pour ce rapport, nous ne sommes pas parvenus à dégager un diagnostic au demeurant, le groupe des rédacteurs se trouve hors de son domaine de compétence sur les questions d'organisation. Nous préférons donc signaler à la fois que nous partageons cette impression négative et que nous ne parvenons pas à étayer ni à analyser ces questions en détail.

Comme il s'agit là d'une question d'importance, nous espérons qu'elle est regardée avec transparence en d'autres lieux.

Nous formulons cependant ci-dessous quelques observations et recommandations qui nous paraissent pertinentes.

## LA GESTION DE CRISE DANS LE MONDE INDUSTRIEL ?

Dans le reste de cette section, nous rapportons des éléments concernant organisation et outillage, en usage dans certains grands secteurs industriels critiques. Pour ce dernier volet, notre conviction est solidement étayée.

### *De l'efficacité à la résilience*

L'optimisation des processus dans un souci d'efficacité croissante introduit, par nature, des spécialisations et rigidités dans les organisations, et donc une diminution des capacités de souplesse d'absorption des chocs et de résilience face aux crises toujours plus variées : les crises récurrentes sont simplement inscrites dans les processus. Cependant, face à des crises imprévues, mais inéluctables, les bénéfices apparents de l'optimisation sont dépassés par les coûts et risques vitaux engagés. Sur le plan plus spécifique de la préparation aux risques, il y a là une grande opportunité de bénéficier des évolutions réalisées dans les vingt dernières années dans les grandes industries : privilégier plus la **résilience** au regard du facteur dominant qu'est la productivité.

C'est GE qui a inventé le terme de *risk officer* (CRO, *Chief Risk Officer*<sup>25</sup>), pour sa filiale qui gérait les crédits, mais ce concept a essaimé dans les autres secteurs industriels. Son rôle est de mettre au point la méthodologie d'analyse des risques au sein de l'entreprise et de faire prévaloir une culture du risque à tous les niveaux. Il est essentiel d'outiller cette activité pour en assurer la flexibilité et l'adaptabilité, faute de quoi la préparation se résume typiquement en la

25 Voir [https://en.wikipedia.org/wiki/Chief\\_risk\\_officer](https://en.wikipedia.org/wiki/Chief_risk_officer)

définition d'une liste de plans spécifiques, prévus dans le menu pour des scénarios imaginés à l'avance, laissant ainsi désarmée l'organisation qui ferait face à une crise inédite.

De façon générale, les trois piliers de la gestion de risque sont : 1/ l'organisation, 2/ la gestion des comportements, et 3/ la modélisation quantitative prévisionnelle concernant les risques.

Tout ceci s'applique pleinement aux gouvernements dans leur gestion de la crise du covid-19, et le restera sans doute pour la gestion des crises à venir.

### **Plates-formes pour gérer le cycle de vie**

Les grands industriels ont développé des grands systèmes complexes critiques (aéronautique et espace et plus largement systèmes de transport, infrastructures de communication, fourniture d'énergie...). Ces systèmes doivent être conçus, fabriqués, puis opérés (puis démantelés le cas échéant), étapes dont l'ensemble constitue **le cycle de vie**.

Penchons-nous sur leur conception. Divers aspects se rencontrent dans la conception de ces systèmes : les fonctions et services qu'ils doivent rendre leur sûreté, leur sécurité, leur résilience leur gestion énergétique leur fabrication. D'autres aspects sont à prendre en compte, selon les secteurs. Ces aspects sont reliés entre eux et sont tous importants. Lors de la conception du système, ces différents aspects sont pris en charge par des équipes différentes (éventuellement chez des sous-traitants), s'appuyant sur des compétences différentes et utilisant des outils numériques différents.

Selon les processus traditionnels, la coordination était réalisée par la remontée des informations et leur échange sous forme synthétisée, selon une organisation arborescente et un fonctionnement en silo, avec le recours à des modélisations qui ne se parlaient pas, sauf à travers l'équivalent des « notes de synthèse », c'est-à-dire les **cahiers des charges**, très mal formalisés. Cette organisation et ce fonctionnement donnaient lieu à des imprécisions et incompréhensions, qui, souvent, se sont traduites dans des erreurs ou défauts découverts tardivement, avec des retards et coûts significatifs<sup>26</sup>.

Pour combattre ces difficultés, selon les évolutions récentes (et encore en cours), ces différents aspects sont regroupés au sein d'une plate-forme unique<sup>27</sup> accompagnant l'ensemble du cycle de vie, et accédant à une base commune et cohérente de données et de modèles (le *PLM, Product Lifecycle Management*<sup>28</sup>). Ces plates-formes permettent un travail collaboratif structuré. À chaque intervenant dans le cycle de vie du système est présentée une vue partielle du contenu de la plate-forme, qui correspond à son rôle et à ses droits d'accès, et lui permet d'apporter, dans la cohérence, sa contribution à l'ensemble. La dernière décennie et

26 Voir par exemple <https://doi.org/10.1561/10000000053> et particulièrement le chapitre introductif.

27 En particulier, deux grands industriels européens du numérique sont fortement présents dans cette activité : Dassault-Systèmes (avec la plate-forme 3D EXPERIENCE qui unifie quatre composantes majeures : modélisation 3D et système, simulation multiphysique et multi-échelle, big data & IA, et environnement de collaboration), et Siemens avec l'outil SIMCENTER focalisé sur la simulation numérique et multiphysique.

28 Voir [https://fr.wikipedia.org/wiki/Gestion\\_du\\_cycle\\_de\\_vie\\_des\\_produits](https://fr.wikipedia.org/wiki/Gestion_du_cycle_de_vie_des_produits)

la pression de la transition environnementale conduisent le secteur industriel des transports (et particulièrement l'aéronautique) à développer des plates-formes dites MDAO (*Multidisciplinary Design, Analysis, and Optimization*) destinée à explorer les grands compromis coût/service rendu/consommation énergétique en phase très amont de la conception de systèmes futurs<sup>29</sup>

**Notre conviction est qu'on est mieux préparé si l'on dispose d'une plate-forme, infrastructure apte à accueillir des familles diverses de modélisations et les données correspondantes. De telles plates-formes peuvent accueillir, rapidement, les modélisations nouvelles qui seraient requises pour une crise inédite.**

Un autre secteur d'activité a vu le développement de plates-formes intéressantes : celui de la sécurité, avec des applications dans les domaines militaire et civil<sup>30</sup> Bien des caractéristiques déjà analysées ci-dessus s'y retrouvent, concentrées sur l'opération (ici, la gestion de situations de sécurité). Modèles et données y sont hébergés, avec des rôles prédominants pour l'analyse de situation et la logistique.

De façon non surprenante, des industriels de ces deux secteurs se sont positionnés sur la gestion de la crise du de la covid-19, voir les entretiens en annexe de ce rapport.

**Au total, les grands industriels semblent beaucoup mieux organisés et préparés en termes de gestion de projet et de gestion de crise que notre pays.**

## QUELQUES OBSERVATIONS SUR LES ACTIONS DE L'EUROPE

**D**ans ce tour d'horizon, la présence des organisations européennes doit être rappelée. Les négociations de l'Europe sur la disponibilité de vaccins et les financements de la crise ont été déterminantes. On a pu noter la mise en place d'outils pour soutenir la recherche, l'économie, les transports, et même combattre la désinformation<sup>31</sup>.

En ce qui concerne la recherche, on peut noter :

- le lancement d'appels à projet spécifiques covid<sup>32</sup>, dans lesquels l'Inserm, par exemple, joue un rôle de pilote ;
- L'auto-mobilisation des Académies des sciences et des technologies européennes Euro-

29 La NASA promeut ce concept qu'elle a contribué à développer (<https://rdcu.be/bpbhm>), et Airbus s'est doté d'une plate-forme de ce type, toujours en cours d'enrichissement (<https://hal.archives-ouvertes.fr/hal-02335530>).

30 Un exemple présenté en annexe de ce rapport est la plate-forme SDP (Security Digital Platform) de Thales.

31 <https://www.consilium.europa.eu/fr/policies/coronavirus/covid-19-research-and-vaccines/>

32 <https://www.horizon2020.gouv.fr/cid150741/appels-a-projets-en-lien-avec-le-covid-19.html>

CAS<sup>33</sup> allemande<sup>34</sup>, italienne<sup>35</sup>, néerlandaise<sup>36</sup>, a été très variable, traduisant sans doute des moyens et pratiques fort disparates.

On peut considérer qu'il est dommage que ces moyens de réflexion technologique n'aient pas été plus mobilisés par les pouvoirs publics.

Une initiative récente est à noter : l'appel à projets entre la France et l'Allemagne sur les technologies d'intelligence artificielle pour la prévention des risques, la gestion des crises et la résilience<sup>37</sup>, publié en février 2021 et ouvert jusqu'en avril 2021. Selon le libellé même de cet appel, il s'agit de soutenir des projets bilatéraux sur le développement de technologies d'IA permettant des avancées significatives sur la prévention des crises en matière de :

- **santé** : détection précoce des épidémies, compréhension de la propagation des virus au sein d'une infrastructure, simulations de stratégies de confinement et de déconfinement, tests virtuels de vaccins, etc.
- **économie** : anticiper les impacts des mesures de confinement et optimiser leur mise en œuvre, automatiser l'appareil productif, accompagner la transformation temporaire des lignes de production, sécuriser et optimiser la chaîne logistique, maintenir le lien à distance avec le client, etc.
- **environnement** : détecter et prévenir les crises environnementales, monitoring des émissions de CO<sub>2</sub>, préservation de la biodiversité, lutte contre l'artificialisation des sols, utilisation raisonnée des intrants en agriculture, etc.»

33 <https://www.euro-case.org/about-us/mission-and-governance/>

34 <https://www.acatech.de/themen/corona-krise/>

35 <https://www.unibo.it/en/research/popular-science/academy-of-sciences>

36 <https://www.euro-case.org/about-us/member-academies/member-academies-netherlands/>

37 <https://www.bpifrance.fr/A-la-une/Appels-a-projets-concours/Appel-a-projets-entre-la-France-et-l-Allemagne-sur-les-technologies-d-intelligence-artificielle-pour-la-prevention-des-risques-la-gestion-des-criSES-et-la-resilience-51532>

## Chapitre II

### IMPORTANCE DE LA MODÉLISATION

**E**n quelques semaines, la visibilité de la science épidémiologique par la société civile est passée de marginale à majeure.

La disponibilité de données publiques a permis à de nombreux ingénieurs, statisticiens et *data scientists* de faire preuve de créativité dans l'interprétation de ces données et, surtout, de partager leurs vues parfois très différentes de la doxa officielle.

Les travaux de modélisation épidémiologique sont apparus comme étant des outils clés, et ce, aux yeux des pouvoirs publics, des médias, et de la population plus généralement. Il s'agit là d'une évolution sociétale majeure : jamais, sans doute, l'importance de travaux de modélisation dans des secteurs non techniques n'avait été perçue si largement.

L'évaluation de l'impact sur l'économie (dont la modélisation reste probablement largement inconnue du public) a concentré aussi beaucoup d'attentions.

**J**amais, sans doute, l'importance de travaux de modélisation dans des secteurs non techniques n'avait été perçue si largement.

Il s'agit là d'une évolution sociétale majeure.

### LES SERVICES PRATIQUES RENDUS PAR LA MODÉLISATION SONT PRÉCIEUX

**D'**emblée, l'irruption, pendant l'hiver 2020, de la modélisation épidémiologique a eu un impact profond :

- mise en évidence d'indices (« chiffres clé » selon la taxonomie de l'application *TousAnti-Covid*) permettant de synthétiser l'état de la pandémie et d'anticiper son évolution ; ces indices sont utilisés pour étayer les décisions et les argumenter ;
- mise en évidence de la dynamique de l'épidémie avec plusieurs caractéristiques non présentes auparavant par les non-spécialistes :

- croissance/décroissance caractérisée par le facteur dit « R effectif »,
- délai de doublement : le contrôle d'une épidémie qui double d'intensité tous les trois jours n'est pas identique à celui d'une épidémie qui double d'intensité tous les trente jours : le temps permet des mises en œuvre correctrices ;
- retard de deux à trois semaines entre des événements sources (confinement, fêtes...) et leurs conséquences. On a ainsi compris qu'un décalage d'une semaine dans les décisions de confinement de l'hiver 2020 a pu avoir des effets considérables sur le développement de l'épidémie.

La campagne de vaccination, qui porte les espoirs des populations et gouvernements et va requérir des prouesses en termes d'organisation et de logistique, pose un nouveau défi à la modélisation épidémiologique. À ce titre, le cas d'Israël en cet hiver 2021 est intéressant. Au 15 janvier 2021, près de 30 % de la population était vaccinée (ce qui mettait déjà le pays en pointe). Cependant, l'épidémie n'était aucunement jugulée ni même ralentie à cette date<sup>38</sup>.

Cela indique le besoin d'étendre la modélisation de la dynamique de la pandémie, à celle de la pandémie contrôlée par le vaccin.

Examinons de plus près de quoi on parle lorsque l'on évoque la modélisation.

## RAPPEL SUR LES NOTIONS DE SYSTÈME ET DE MODÉLISATION

La notion de système est très vaste et couvre de nombreux domaines : une cellule vivante peut être considérée comme un système, une fusée est un système, la France est un système, la Terre, etc. notre propos n'est pas ici de définir cette notion nous nous contenterons d'en dessiner quelques traits importants. Un système est caractérisé par une enveloppe (sa frontière), des états internes et un environnement, le tout étant fonction du temps : le système change d'état en fonction de son fonctionnement interne et des sollicitations auxquelles le soumet son environnement. La notion plus récente de systèmes de systèmes avec sa notion sœur de système ouvert a quelque peu bousculé cette vision un peu simple. Il existe tout un arsenal mathématique pour modéliser les systèmes et les construire comme assemblages de composants, les contrôler, analyser leur robustesse face aux perturbations, etc. Le pilotage des grands systèmes complexes est une grande réussite de l'ingénierie du xx<sup>e</sup> siècle : gestion du réseau ferroviaire, gestion du système électrique, régulation du trafic aérien, etc.

C'est sous cet angle que nous avons abordé cette étude sur la crise de la covid-19, qui perturbe de façon majeure tout le système socioindustriel français :

- quelles sont les frontières du système à considérer dans le cas de la crise de la covid-19 ? (le système médical à lui seul est insuffisant) ?
- quelles sont les modalités d'observations sur le système dont on dispose ?
- quels sont les outils dont on dispose pour comprendre le fonctionnement du système et analyser les actions possibles et leurs effets ?

38 Voir <https://www.coronavirus-statistiques.com/stats-pays/coronavirus-nombre-de-cas-en-israel/>

- quels outils pour aider les organisations humaines à gérer cette crise ?

La modélisation a pour but la compréhension des systèmes, de leurs évolutions et des possibilités de contrôle et/ou d'optimisation. Un modèle est une description d'un système dans un langage. On peut identifier trois étapes de la modélisation :

- **le modèle discursif et phénoménologique.** Il s'agit de décrire l'objet et son comportement en langage naturel<sup>39</sup>. Ce qu'on peut voir, observer, déduire, les leviers d'actions, etc. Le modèle discursif est l'outil de communication compréhensible par l'ensemble de la population ;
- **le modèle mathématique :** le modèle mathématique correspond à la description d'un ensemble de variables, fonctions et mécanismes mathématiques décrivant le système en langage mathématique<sup>40</sup>. Le modèle mathématique permet de comprendre les difficultés intrinsèques du système. Plusieurs facettes d'un système peuvent être décrites par des modèles mathématiques différents. Le modèle mathématique requiert une qualification importante pour être développé et compris. On peut concevoir plusieurs modèles mathématiques pour un même système, selon les simplifications considérées ;
- **le modèle numérique :** la connaissance des valeurs des paramètres et l'implémentation du modèle mathématique sous forme de code numérique permet de réaliser de nombreux calculs et d'en exploiter les résultats. Le code numérique permet de maîtriser la complexité liée au nombre et à la diversité des variables et des liens entre celles-ci (équations, tables...), la difficulté des calculs, et permet de développer des scénarii (*what-if*<sup>41</sup>) et analyses de futurs possibles en fonction des décisions envisagées. Le code numérique est développé par des spécialistes, mais peut ensuite être mis entre les mains d'une population importante d'acteurs et de décideurs. Il tient ainsi un rôle clef de communication pour le pilotage de systèmes complexes et en gestion de crise. La collecte et l'interprétation des données issues du monde réel sont clefs dans cette étape.

Ces trois niveaux de modèles sont évidemment profondément liés. La **qualité d'un modèle** est évaluée par sa précision de représentation du passé et de modélisation du futur. Il repose donc sur les **éléments de terrain et de réalité**, par le biais de **données** qui sont collectées et confrontées au modèle dans un contexte expérimental en premier lieu, puis utilisées pour l'aide à la décision ensuite.

Le présent rapport n'a pas pour objet de faire le tour d'horizon complet des différentes approches de modélisation de la pandémie de la covid-19. Nous illustrons les aspects clés de ces modèles qui nous ont paru les plus pertinents à comprendre dans ce contexte complexe de la gestion de crise de la covid-19.

Quelques pointeurs intéressants sont fournis en annexe de ce rapport. Néanmoins, pour

39 Par exemple, le modèle discursif de la gravitation consiste à dire : 1-> Tous les corps s'attirent », 2-> Plus les corps sont lourds, plus ils attirent les autres », 3-> Plus les corps sont loin, moins ils s'attirent ». Le modèle discursif ne permet pas de savoir si un corps lourd et lointain attire plus ou moins qu'un objet léger et proche.

40 Le modèle mathématique de la gravitation va utiliser par exemple l'équation de Newton  $F = GM1M2/d^2$ . Ce modèle permet de prévoir les formes elliptiques des trajectoires. Il ne suffit pas à faire des calculs précis sans la connaissances des paramètres physiques (masse des planètes, distances, etc).

41 Cf Chap.IV, § Aide à la décision

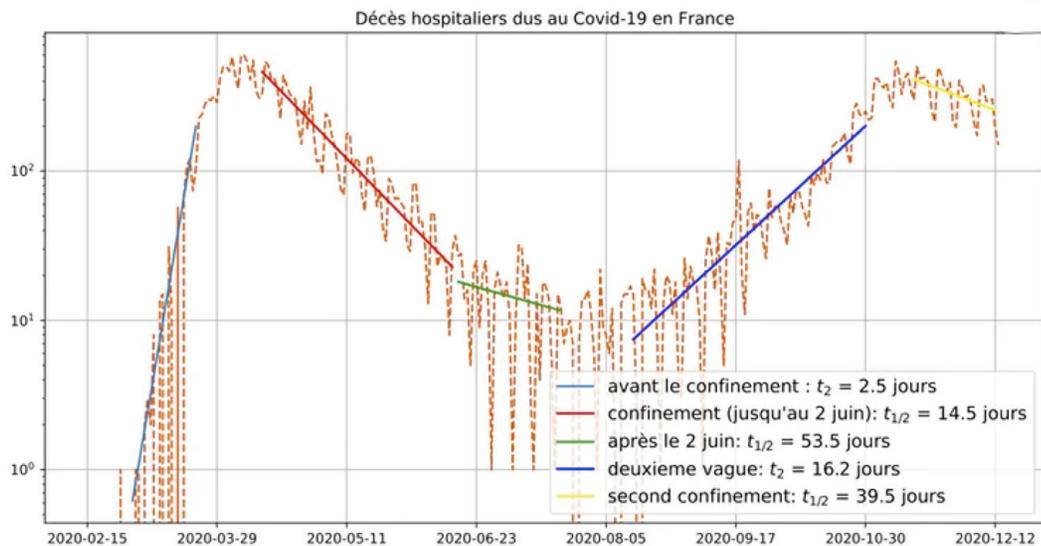
préparer le lecteur aux difficultés du sujet et motiver encore plus la section qui suit, nous invitons tout particulièrement le lecteur à lire l'article ci-référence<sup>42</sup> publié dans le *Quanta Magazine*.

## DOMAINES DE MODÉLISATION CONSIDÉRÉS DANS LE CADRE DE LA COVID-19

Dans cette section nous passons en revue la grande variété des approches de modélisation, en commençant par l'épidémiologie puis en passant en revue d'autres domaines (tout en privilégiant ceux qui concernent le covid). Les principaux domaines utilisés et que nous avons pu recenser sont les suivants :

### Modélisation de l'épidémie au niveau d'une population

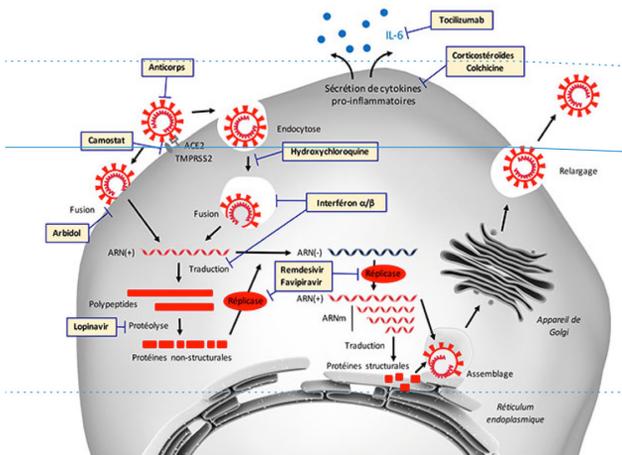
On met dans cette catégorie tous les modèles calculant le nombre de malades, décès,



Il convient également de mentionner certaines études destinées à clarifier si oui ou non la température et l'humidité ambiantes ont une influence sur la propagation du virus, et combien.<sup>43</sup>

### Modélisation des cellules et virus

Il s'agit d'une modélisation de nature biologique, depuis les structures atomiques jusqu'aux structures des virus et cellules, incluant l'ADN. On peut, par exemple, modéliser les divers chemins d'accès et impacts du virus dans les cellules, reconstituer (*in*



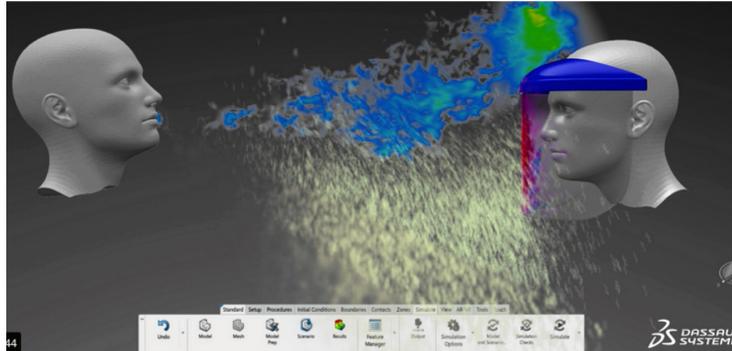
[3-hard-lessons-of-modeling-the-coronavirus-pandemic-20210128/#rect.com/science/article/pii/S1279847921000021?via%3Dihub](https://www.nature.com/articles/s41586-021-00021-7)

*in vitro*) le complexe de réplication du de la covid-19 et le modéliser *in silico*.

Un tel exemple de modélisation pour l'action d'un médicament est présenté ci-contre (Croquis : source Univ. McGill, Canada).

### Modélisation des mécanismes physiques de transmission

On place dans cette catégorie notamment les simulations aérodynamiques de transport des virus en espace clos/ouvert : (source Dassault Systèmes).



### Ingénierie biologique

On peut aussi rappeler l'importance, comprise récemment, de l'ingénierie biologique avec l'arrivée de vaccins à ARN messager. Caractériser un virus par son ADN, puis cartographier un ARN de réponse, est une forme de modélisation<sup>44</sup>

## PANDÉMIES HUMAINES : EN QUOI CELA CHANGE LA DÉMARCHÉ DE MODÉLISATION

Il serait bien présomptueux de prétendre modéliser d'emblée et en anticipation tous les facteurs importants qui interviennent dans une crise inédite : on en fait précisément l'identification lors de l'apparition de la crise, au fur et à mesure de son développement.

Cependant, certains volets demeurent des invariants : s'agissant des crises sanitaires, et donc, de pandémies humaines, on trouvera toujours l'épidémiologie, la logistique, l'état du système de santé, et, plus globalement, l'état d'avancement des solutions médicales. Sous le vocable de « système de santé », il convient d'inclure l'industrie de santé (et peut-être de la recherche) dans ses composantes nationale et internationale.

En préliminaire, une interaction semble aujourd'hui confirmée entre situation météorologique et propagation du virus dans la pandémie<sup>45</sup>. Mais, au-delà de l'interaction entre ces deux facteurs sur lesquels nous n'avons guère prise, l'important est de comprendre les différentes facettes et composantes du système.

////////////////////////////////////

44 Voir l'excellent article explicatif <https://renaudguerin.net/posts/explorons-le-code-source-du-vaccin-bio-tech-pfizer-sars-cov-2/>

45 <https://www.francebleu.fr/infos/sante-sciences/coronavirus-la-societe-montpellieraine-predict-fait-un-lien-precis-entre-la-pandemie-et-la-meteo-1608307346>

## Psychologie et intelligence des individus, anticipations de la population

Comment prendre en compte le comportement des individus d'une population en réaction à

Les humains ne sont ni des plantes ni des animaux, mais des êtres sociaux : quand un gouvernement annonce une décision, la population s'adapte à la décision, et fait éventuellement un peu autre chose.

la pandémie ou aux mesures édictées. Les humains ne sont ni des plantes ni des animaux, mais des êtres sociaux : quand un gouvernement annonce une décision, la population s'adapte à la décision... et fait éventuellement un peu autre chose. Les économistes sont habitués à ce fait depuis la « critique de Lucas<sup>46</sup> ». Le concept d'anticipation rationnelle en a résulté, avec le pendant mathématique des équilibres de Nash<sup>47</sup>

en théorie des jeux. 98 % des modèles généraux développés en épidémiologie sont appliqués aussi bien aux humains qu'aux animaux ou végétaux, or les humains ne réagissent pas comme les animaux (c'est le sens de la Critique de Lucas). De plus, les choix effectués par l'individu sont souvent antagonistes du bien collectif. En ce sens, il y a une insuffisance méthodologique fondamentale dans les approches actuelles en épidémiologie. Voir à ce sujet le développement sur les MFG (Mean Field Games) (chapitre IV), qui apportent une réponse mathématique intéressante.

## Industrie et logistique

Les mois de mars et avril 2020 resteront en mémoire pour ce qui est des défis logistiques et industriels rencontrés, particulièrement en France : masques, respirateurs et tests, mais aussi simple approvisionnement en biens de base pour la population. En ce début 2021, la campagne mondiale de vaccination constitue un formidable défi industriel (urgence de production) et logistique<sup>48</sup>.

## Économie

L'impact de la pandémie sur l'économie française est considérable et sera durable<sup>49,50</sup>. Et il en va de même pour le monde dans son ensemble, à l'exception notable de la Chine, qui s'en tire mieux<sup>51</sup>. La cause en est l'ensemble des mesures de confinement. L'impact est variable selon les secteurs économiques<sup>52</sup>. Au total, la crise a affecté de façon très inégale les différents

46 Robert Lucas Jr, qui est un des plus « grands » prix Nobel d'économie, est connu pour son rôle dans cette évolution fondamentale : les économistes appellent cela « La critique de Lucas ». Cette « critique » a été formulée par Robert Lucas Jr en 1976 et a apporté un changement important dans la façon dont les modèles économétriques sont construits. Elle peut aujourd'hui être transposée aux modèles épidémiologiques. Voir annexe.

47 Cf [https://fr.wikipedia.org/wiki/%C3%89quilibre\\_de\\_Nash](https://fr.wikipedia.org/wiki/%C3%89quilibre_de_Nash)

48 Voir <https://www.nature.com/articles/d41586-020-03134-2>

49 Pour l'impact sur l'économie française, voir [https://www.insee.fr/fr/statistiques/fichier/4997475/note-conj\\_151220.pdf](https://www.insee.fr/fr/statistiques/fichier/4997475/note-conj_151220.pdf)

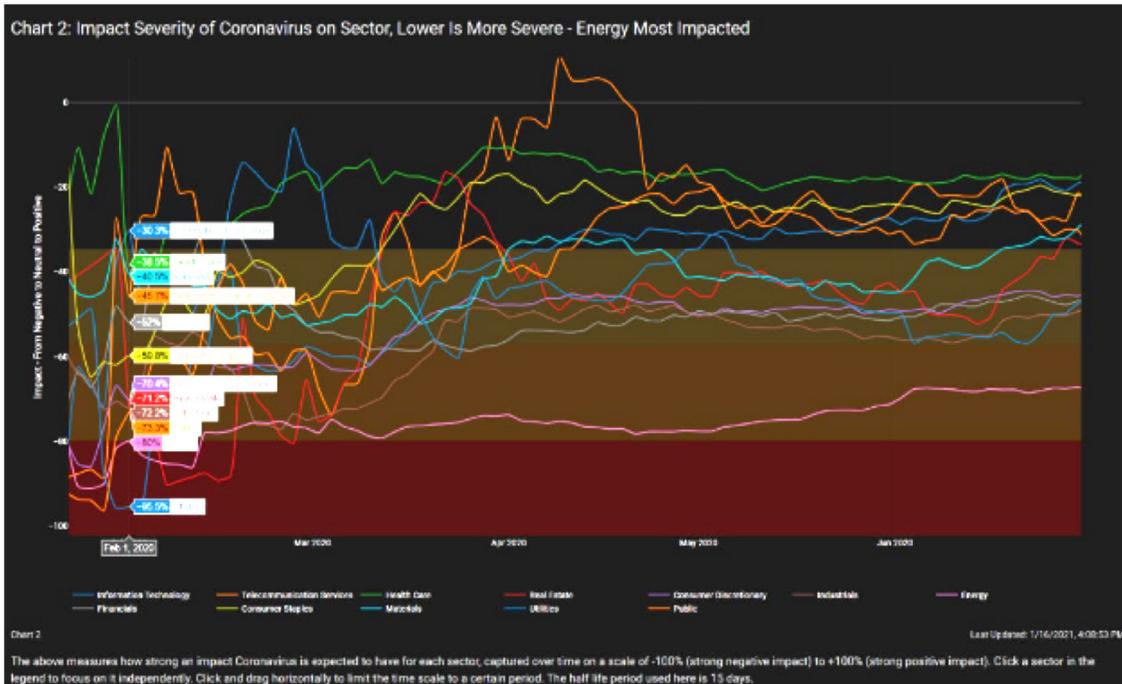
50 Pour l'impact sur l'activité des sociétés en France, voir <https://www.insee.fr/fr/statistiques/4994488>

51 Voir <https://www.bbc.com/news/business-55699971>

52 Pour l'impact sur le comportement des consommateurs français selon le secteur, voir L'idée : [https://www.google.com/search?client=firefox-b-e&q=https%3A%2F%2Fwww.insee.fr%2Ffr%2Fstatistiques%2Ffichier%2F4997475%2Fnote-conj\\_151220.pdf](https://www.google.com/search?client=firefox-b-e&q=https%3A%2F%2Fwww.insee.fr%2Ffr%2Fstatistiques%2Ffichier%2F4997475%2Fnote-conj_151220.pdf)

pays du monde. Les grands absents de cette crise sont des pays majeurs de l'Asie (Chine, Japon, Vietnam, Corée, Taïwan) ainsi que l'Afrique. Si les conséquences économiques sont majeures dans les pays affectés, elles sont le plus souvent mineures dans les autres : la crise de la covid-19 est ainsi une opportunité de prise de parts de marché inhabituelle.

À titre d'illustration, le diagramme suivant rend compte des impacts économiques dans les chaînes de valeur, avec les effets dominos sur les différents secteurs (source *Causality Link*, voir l'annexe 3 pour plus de détails).



### Société et culture de société

Le cas de Taïwan est frappant à cet égard : une étude récente<sup>53</sup> met en évidence l'importance de facteurs sociétaux et d'organisation dans la résilience remarquable de ce pays : prise de conscience très tôt dans l'hiver 2020, préparation culturelle de longue date, y compris par l'expérience acquise lors des épidémies antérieures, efficacité du traçage et, plus généralement, de l'organisation. La même constatation vaut, avec des résultats moins spectaculaires, pour la Corée du Sud<sup>54, 55</sup>, et, malgré les incertitudes qui affectent les chiffres officiels, pour la Chine<sup>56</sup>. Un facteur important a été et demeure la « contrôlabilité » (pour utiliser le langage imagé des automaticiens) de la population quant à son respect des règles sanitaires imposées et le degré d'acceptabilité de celles-ci : tout ceci a varié selon les pays et leurs traditions. Il en va de même pour les effets secondaires sanitaires de la pandémie (psychologiques et, aussi, le suivi des maladies autres

53 Voir <https://www.tandfonline.com/doi/full/10.1080/02185385.2020.1819396>

54 Voir [https://www.eurekalert.org/pub\\_releases/2020-08/uocd-wtr081020.php](https://www.eurekalert.org/pub_releases/2020-08/uocd-wtr081020.php)

55 Voir <https://www.korea.net/NewsFocus/Society/view?articleId=185127>

56 Voir [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30800-8/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30800-8/fulltext)

que la covid-19). Pour un même virus<sup>57</sup> et des caractéristiques de population similaires, le développement de la pandémie a varié de manière importante selon les régions du monde.

### *L'expérience acquise*

L'expérience acquise sur les crises passées se doit d'être archivée pour la capitalisation des connaissances. Se rappeler le passé des situations de crise, les solutions qui marchent et qui ne marchent pas est une nécessité pour ne pas redécouvrir et devoir tout réinventer dans l'urgence.

Nous avons pu constater que les pays asiatiques avaient tiré les leçons sur l'organisation

**A**fter MERS, the country made a series of policy changes to improve pandemic preparedness and response. When COVID-19 struck, the painful memory of MERS inspired an early, aggressive government response — and a willingness among people to wear masks, cooperate with contact tracers, and otherwise listen to public health officials. For example, wearing a mask in public spaces, already common because of air pollution, became a social norm early in the pandemic. A recent poll showed that more people adhered to public prevention protocols during the COVID-19 outbreak than during the MERS outbreak.

et la gestion des crises de santé précédentes (H1N1, MERS...). En particulier, la Corée du Sud a tiré profit de son expérience de la crise du MERS, comme illustré par l'encadré ci-contre, extrait de la référence<sup>58</sup>. Pour plus de détails, voir le paragraphe sur la Corée dans l'annexe bibliographique, section 4.

Pour la France des enseignements ont été tirés et compilés, mais il semble qu'ils n'aient pas été utilisés, ni en préventif, ni pour la gestion (cf. le rapport du Sénat<sup>59</sup>).

### *Psychologie et modélisation*

**D**ans la gestion de crise, le facteur humain est extrêmement important. Le suivi psychologique a montré des évolutions inquiétantes du nombre de dépressions et des suicides dans cette période covid, liées notamment aux difficultés économiques, à l'isolement, à l'ennui, au manque d'activité physique, appétence à suivre les recommandations du gouvernement, etc. L'enquête nationale COVIPREV<sup>60</sup> permet une forme de suivi dans plusieurs dimensions de ces paramètres. Le suivi psychologique et comportemental quantitatif est très intéressant :

- pour les aspects de santé : quels comportements influencent la propagation de l'épidémie ? Développer des outils pour mieux évaluer cela est aussi un moyen d'améliorer les outils mathématiques et informatiques de prévision, ainsi que les outils de communication ;
- pour les aspects économiques : il permet de suivre et comprendre des transformations économiques majeures : les commandes de biens par ordinateur, le télétravail, et même

57 Nous faisons référence à la période qui a précédé l'émergence de mutations significatives du virus.

58 Emerging covid-19 success story: South Korea learned the lessons of MERS. Voir par exemple <https://ourworldindata.org/covid-exemplar-south-korea>

59 [https://www.senat.fr/commission/enquete/gestion\\_de\\_la\\_crise\\_sanitaire.html](https://www.senat.fr/commission/enquete/gestion_de_la_crise_sanitaire.html)

60 <https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19>

des migrations d'habitation principale, sont des transformations des grands équilibres macroéconomiques. Par exemple : le confinement développe mécaniquement le téléachat. Combien ? Par ailleurs, l'ennui développe-t-il les achats compulsifs ? Comment mesurer cette relation ? etc.

Les modèles issus de la psychologie sont des modèles discursifs et phénoménologiques, qui s'appuient sur des données statistiques, mais qui ne permettent pas encore l'élaboration d'outils mathématiques et informatiques connectés de façon simple aux autres modèles (santé, économiques, etc.) que nous avons pu trouver par ailleurs. Ainsi, modéliser numériquement la nature humaine reste une immense difficulté. Aboutir à des approximations statistiques de comportements serait néanmoins intéressant et utile pour le reste de la modélisation.



## Chapitre III

# DES DONNÉES À LA MODÉLISATION ET À LA MISE EN ŒUVRE NUMÉRIQUE

**C**es trois composantes sont les piliers indissociables d'une modélisation réussie.

### LA QUESTION DES DONNÉES

Une des grandes leçons de cette crise est que « les données ne sont pas données » : les paramètres clef ne sont pas évidents les données sont difficiles à collecter et valider elles sont difficiles à mesurer elles sont difficiles à interpréter enfin, le temps (avec l'obsolescence qui en résulte) est un paramètre important.

Une des grandes leçons de cette crise est que « les données ne sont pas données »

#### *L'identification des bons paramètres à suivre*

La crise de la covid 19 a rapidement démontré que le simple suivi des décès dans la pandémie était très insuffisant, car cet indicateur arrivait très tard dans le processus. D'autres indicateurs plus avancés (entrées en réanimation, entrées aux urgences, cas déclarés) ont été mis en place qui permettent d'anticiper les décès de 8 à 15 jours, et donc d'adapter le système de soins. Ces indicateurs ne sont pas les seuls à être devenus pertinents. D'autres paramètres, indicateurs des comportements et de la crise économique sont apparus progressivement :

- **Suivi des mobilités** (Google, TomTom<sup>61</sup>), trafic automobile (Sytadin<sup>62</sup>), transports publics (trafic RATP, SNCF), etc. : le graphique ci-joint regroupe des données collectées via les smartphones et permettant de mesurer l'écart par rapport à la moyenne pour des comportements de transport, de déplacements, de télétravail. Les indicateurs de trafic ont le

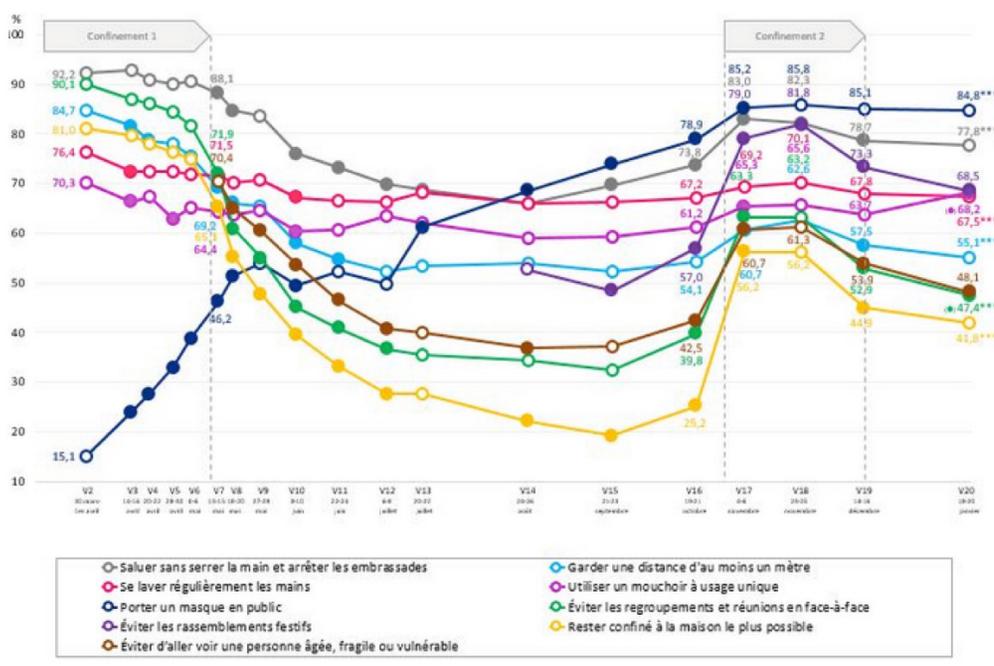
61 Cf <https://www.tomtom.com/blog/moving-world/covid-19-traffic-mobility/>

62 [http://www.sytadin.fr/sys/barometres\\_de\\_la\\_circulation.jsp.html#](http://www.sytadin.fr/sys/barometres_de_la_circulation.jsp.html#)



- **Suivi des comportements** (cf. Santé publique<sup>63</sup>) : il s'agit d'un système d'enquêtes répétées sous forme de questionnaire régulier avec un échantillon de 2 000 personnes. Ceci permet de suivre les déterminants les plus susceptibles d'affecter à court terme la situation sanitaire, notamment l'adoption des mesures de protection (déterminants de la contagion) et la santé mentale, également susceptibles de présenter un fardeau supplémentaire pour le système de santé.

### Fréquences de l'adoption systématique déclarée des mesures de prévention et évolutions (% pondérés). Enquête CoviPrev, France métropolitaine, janvier 2021



- **Indicateurs indirects de propagation du virus** : suivi des eaux usées<sup>64</sup>. Cette méthode de mesure, lancée par les marins pompiers de Marseille a été un grand succès<sup>65</sup> et a débouché sur la création du réseau COMETE<sup>66</sup>. En dehors de l'aspect quantitatif global d'une ville par des tests simples et rapides, le suivi qualitatif (détection et présence des variants du virus) est aussi un axe prometteur d'utilisation de ces outils. Ces indicateurs donnent en temps réel (une semaine avant l'évolution du nombre de cas recensés selon les marins-pompiers de Marseille) une mesure globale de la diffusion du virus dans la population. Le processus de diagnostic utilisé permet d'aboutir à une précision progressive du trajet des eaux usées

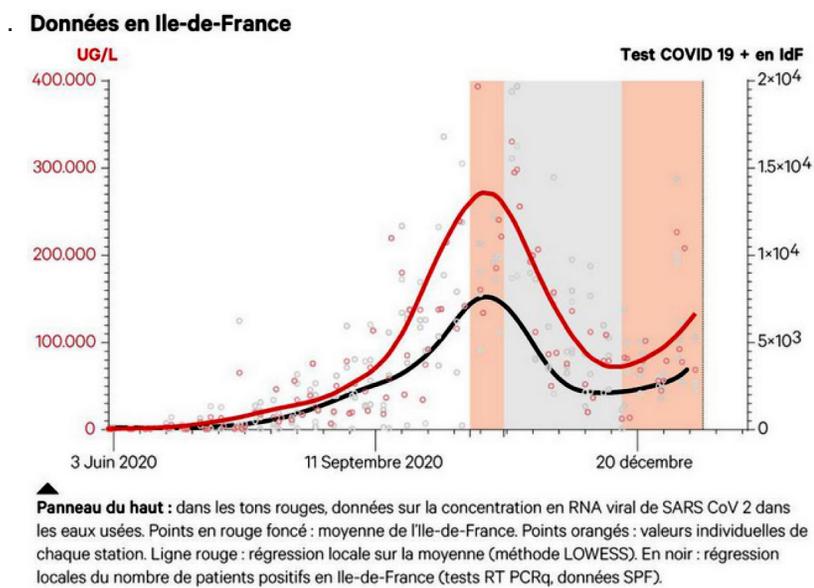
63 <https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19>

64 <https://www.lesechos.fr/industrie-services/energie-environnement/covid-les-eaux-usees-revelent-une-inquietante-montee-du-virus-1281066>

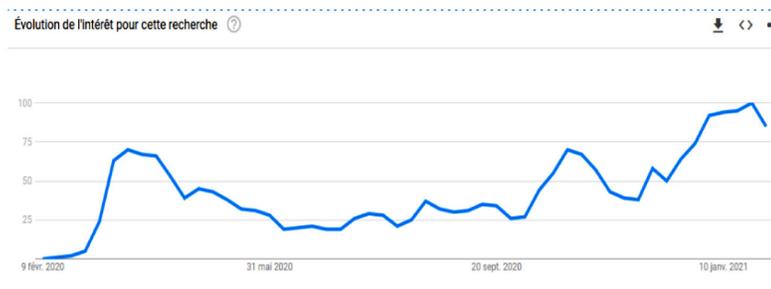
65 <https://madeinmarseille.net/82388-lanalyse-des-eaux-usees-cartographiee-pour-anticiper-levolution-de-la-covid-dans-la-region/>

66 Cf entretien réseau COMETE : P Guérin

via les canalisations jusqu'au bâtiment et ensuite au logement par les analyses de surface, avec une très grande réactivité. Le réseau de recherche Obépine, présent sur l'ensemble du territoire, permet une mise à jour deux fois par semaine des données. Il faut noter que les systèmes mis en place disposent à la fois d'une grande sensibilité, et sont calibrables par rapport à des dilutions (eaux de pluie...)



- **Indicateurs de recherche de mots-clés dans les moteurs de recherche.** Ils se corrént assez bien avec le taux de décès : Google avait déployé dans le passé le service Google Flu, qui permettait de prédire l'intensité des crises de grippe, mais l'avait retiré face à la faible précision quantitative atteinte. Cependant, si une grande précision n'est pas exigée, cet indicateur, comme les eaux usées, fournit des évaluations intéressantes sur les tendances de propagation virale.



### Collecte

**Le coût de collecte et d'organisation de ces données** et indicateurs est très variable. Ces données peuvent résulter de traitements de routine pour des entreprises de gestion de données (opérateurs télécoms, Google...), de collectes spécifiques (données réelles de transport public), parfois semi-manuelles (données des hôpitaux et praticiens de santé). Si beaucoup de données agrégées ont été cédées gratuitement aux pouvoirs publics, on peut aussi envisager que le coût de mise à disposition des informations aux pouvoirs publics soit compensé financièrement.

La collecte de ces informations pose des **problèmes techniques d’anonymisation** et de droits. Dans le cadre du **RGPD** se pose la question du consentement de l'utilisateur à un traitement qui n'est pas prévu : l'opérateur peut-il publier des données ? L'une des questions soulevées est celle du stockage de ces données. « *Cela demande des ajustements réglementaires et un accord de la CNIL. Il faudrait pouvoir garder des données sur une durée de temps longue or, actuellement, nous devons les supprimer au bout d'un an* », a reconnu Stéphane Richard, le PDG d'Orange, au [Figaro](#).

La **durée et le retard de collecte** sont aussi des difficultés qui ont été rencontrées : la collecte manuelle des décès covid faiblit considérablement le week-end pour bondir le lundi. Autres soucis de saisie pour les données des Ehpad, qui ont été agrégées plusieurs fois sur des périodes de plusieurs jours vers un seul jour de déclaration, sans oublier que le département des Bouches-du-Rhône n'a pas fourni de données pendant plusieurs semaines en avril 2020.

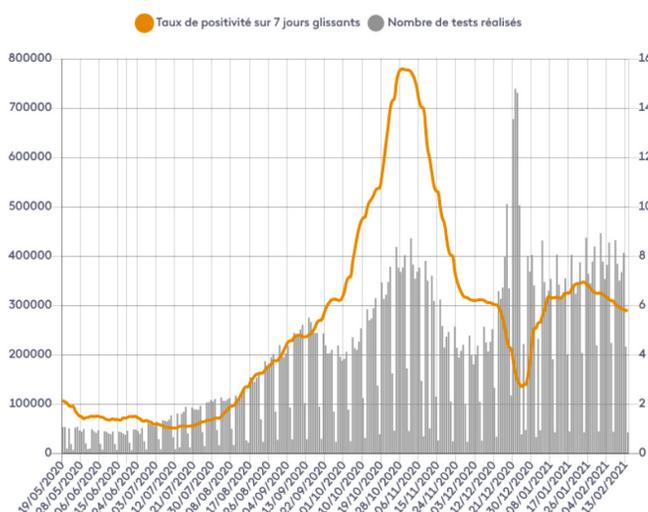
Le coût des tests biologiques PCR est aussi une interrogation, alors que l'on sait que les analyses biologiques de coronavirus à titre vétérinaire chaque année sont très nombreuses et que cette profession dispose de nombreuses machines de séquençement fonctionnant à des tarifs très bas.

### Validation et interprétation

La qualité des données est un vrai sujet. Changer la méthode de comptage en cours de route est une vraie question !

On l'a observé par exemple dans le décompte des cas covid et du taux de positivité<sup>67</sup>:

- en mars/avril les cas décomptés sont essentiellement très proches des arrivées en hôpital ;
- ensuite la mesure prise en compte est le nombre de tests PCR positifs, mais ce nombre dépend fortement du nombre de tests réalisés : le nombre de cas augmente avec nombre de tests ;
- puis le nombre de cas est calculé comme le ratio (nombre de cas positifs/nombre de tests) ;
- enfin ce ratio est modifié le 8 décembre 2020, pour 'prendre en compte les cas de test multiples de personnes se faisant tester plusieurs fois à intervalle rapproché'.



Évidemment, ces variations de méthode introduisent une discontinuité d'interprétation, mais ne sont pas recensées explicitement dans des bases différentes.

////////////////////////////////////

67 Source : <https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/>

Enfin, les données recueillies doivent être interprétées avec précaution. Par exemple, la période de Noël montre une différence de comportement face au test (le nombre de test est multiplié par deux et le taux de positivité est divisé par deux), et non un « miracle de Noël<sup>68</sup> »:

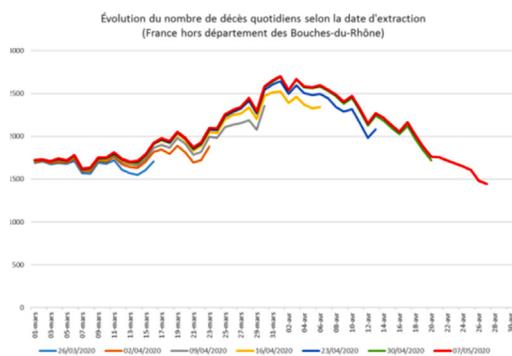
Donnons deux exemples sur la méthodologie de comptage qui n'est pas si simple à définir :

- le débat de savoir si les décès ont tous été **PAR** la covid-19 ou **AVEC** la covid-19 reste entier à ce stade. Le nombre de décès avec des comorbidités élevées est tellement important que la mortalité ajoutée n'est pas toujours évidente à définir : s'il n'y a pas de surmortalité comme en certaines périodes, mais qu'il y a des décès catalogués covid, que doit-on en conclure ?
- la réalisation de campagnes de tests PCR très importantes et coûteuses a permis de remonter des séries de comptage de positivité. Cependant, le nombre de cycles utilisé n'est pas reporté dans les statistiques, alors qu'il a été établi que le taux d'erreur grandit avec le nombre de cycles<sup>69</sup>. On aurait ainsi souhaité (pour les spécialistes en charge des analyses) disposer de résultats quotidiens sous forme de triplets (nombre de cycles PCR, type de test, numéro de la machine), au lieu de résultats binaires « positif/négatif » et de leur simple agrégation globale (nombre de cas détectés « positifs ») et par tranche d'âge. Une occasion de disposer d'une métrique publique précise a été négligée alors qu'elle avait été proposée par le SI-DEP dans sa note du 19 juin 2020<sup>70</sup>.

Certaines données (eaux usées, mobilités) sont entachées de grandes marges d'erreurs, et ne peuvent être utilisées au mieux que comme indicateur en variation, car elles ne peuvent être corrélées de façon fiable au niveau de santé de la population.

Malgré tous leurs défauts, les données ont permis néanmoins de découvrir des informations intéressantes sur les mécanismes de propagation de la maladie. Le décalage entre les pics des différentes courbes renseigne sur les durées de propagation, d'infection et sur la létalité.

### L'importance du temps



L'importance du temps se manifeste sous plusieurs formes. Tout d'abord, la constitution de séries temporelles est indispensable pour obtenir des prévisions encore faut-il les produire à temps !

L'ancienneté ou la fraîcheur des données varie. Ainsi, plusieurs points ont été notés dans cette période covid : les données sont disponibles tardivement, et susceptibles de subir des évolutions au cours du temps. Un exemple simple est fourni par les statistiques de décès en France de l'Insee : la remontée des décès

68 Source idem.

69 <https://www.publichealthontario.ca/-/media/documents/ncov/main/2020/09/cycle-threshold-values-sars-cov2-pcr.pdf?la=fr>

70 Système d'Information national de dépistage populationnel (SI-DEP) , ministère des solidarités et de la santé.

par les mairies prend du temps, il se compte en jours dans 90 % des cas, mais prend des mois voire des années pour être comptabilisée très précisément. La courbe publiée par l'Insee au jour le jour montre ainsi une sous-estimation d'au moins 15 % par rapport à un chiffre plus juste qui arrive avec six semaines de délai.

La disponibilité à temps d'indicateurs avancés est donc un point clé. Si on n'en dispose pas, seule la modélisation fine des erreurs systématiques de mesure permet d'approcher des chiffres réalistes en temps réel.

Enfin, le temps humain intervient : le temps est nécessaire aux esprits pour comprendre les divers mécanismes mis en jeu au cours de l'épidémie, et accumuler les connaissances nécessaires à la modélisation.

### ***Abstraire les données, mais tracer les abstractions***

**D**ans la crise du de la covid-19 la méthode de comptage des personnes infectées a changé trois fois. Des biais variables en ont résulté, qui ont été renforcés par le système de collecte des informations. L'entrée des informations était : + ou -. Et donc d'où venait le + ? L'origine du + ou du - était perdue. Or, la décision finale pouvait dépendre du kit utilisé et/ou de la méthode d'interprétation. Ce problème de biais est récurrent et inévitable avec les synthèses en arborescence où l'on abstrait de plus en plus les informations.

Pourtant, synthétiser et abstraire est nécessaire. Comment résoudre ce dilemme ?

La solution est de rendre cette abstraction réversible par le moyen de la traçabilité : si l'on peut, depuis l'information synthétisée, remonter à l'information brute, alors les processus de décision peuvent, au besoin, être rejoués de façons à gommer les biais variables. Par exemple, en retraçant les données brutes pour reproduire ce qu'aurait été la synthèse avec une nouvelle méthode de décision.

**Tracer est donc un service clé : or, c'est précisément le point fort de ces technologies de plate-forme que de tracer et relier tous les types de données, de modèles, d'algorithmes, et d'information.**

### ***Conclusions sur les données***

**U**n épisode de crise aussi complexe que celle due à la covid-19 a amené à utiliser des données de nature très variées. Il faut reconnaître qu'un effort important et inédit a été fait pour partager les données disponibles.

Il apparaît toutefois que la documentation des différentes méthodes de collecte de données doit être encore renforcée. En particulier, les méthodologies de collecte de données sont à mieux documenter et doivent être accompagnées d'informations sur les incertitudes.

## MODÉLISATION : DOMAINES ET OUTILS MATHÉMATIQUES

On rappelle dans ce paragraphe quelques outils mathématiques permettant d'aborder la compréhension des données et les prévisions. On les présente dans un ordre de complexité croissante :

- modélisation à partir de données d'apprentissage ;
- modèles d'évolution à base d'équations différentielles (phénomènes à temps continu) et agents (phénomènes à temps discret) ;
- théorie de la décision : équilibres de Nash, jeux à champ moyen (MFG), etc.
- et on aimerait inclure dans cette liste les modélisations pour la psychologie...

Les modèles peuvent ou non inclure des aspects probabilistes (bruits et incertitudes, échantillons d'une grande population, défaillances, etc.). Au total, apprentissage et modélisation (avec le calcul scientifique en particulier) figurent aujourd'hui parmi les grands consommateurs de ressources de calcul.

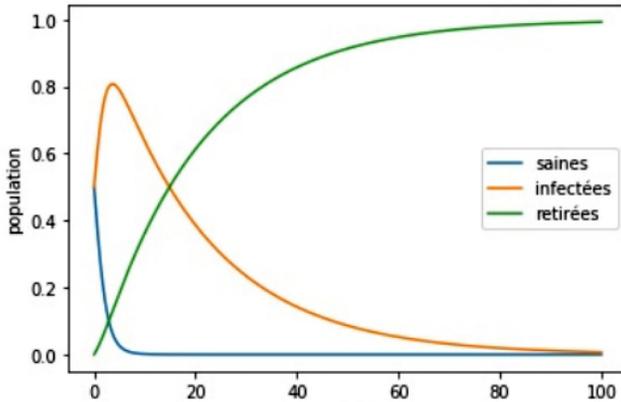
### *Modèles basés sur de l'apprentissage*

Une première classe intéressante est celle des modèles dits *boîte noire* établis directement par exploitation de (nombreuses) données. Il s'agit là de modéliser des comportements pour lesquels on n'a pas de connaissance *a priori* s'appuyant sur des considérations de physique. Par exemple, lorsque des données en nombre suffisant sont disponibles à tout moment, les analyses de séries temporelles, traditionnelles en statistiques, permettent d'effectuer des projections, d'identifier des corrélations ou de détecter des changements. Dans le cas de la covid-19, des techniques de séries temporelles donnent des extrapolations à court terme, mais sont impuissantes à prédire l'effet d'une mesure de contrôle de l'épidémie (confinement...).

Les techniques d'apprentissage (*machine learning*) permettent de construire des modèles boîte noire par l'exploitation d'un grand nombre de données. Ces techniques ont vu une explosion dans la décennie 2010 et ont causé l'irruption au premier plan du sigle-clé IA (intelligence artificielle) dans le grand public, les médias et les mondes industriel et politique. En guise d'exemple bien connu, Google ou Criteo propose aux usagers des publicités *personnalisées*, grâce à des algorithmes de type *recommandation* (si l'individu  $x$  au sein d'une population a exprimé de l'intérêt pour certains types d'information, alors il sera probablement intéressé par telle ou telle publicité  $y$ ). Il s'agit là de modéliser des comportements pour lesquels on n'a pas de connaissance *a priori* s'appuyant sur des considérations de physique ou autre. Les techniques de *machine learning* permettent de construire ce genre de modèle par l'exploitation massive de données (Google a accès à une grande population d'utilisateurs). Malgré les progrès du *machine learning* et tout particulièrement de l'apprentissage profond, l'évaluation prévisionnelle (*What-if*) de politiques non expérimentées auparavant (et donc non porteuses de données) ne semble pas réalisable par ces techniques.

### Modèles de type SIR « classiques » et modèles à base d'agents

Le modèle le plus simple/populaire de propagation d'épidémie est le modèle SIR<sup>71</sup>. On partitionne la population en groupes (saine, Infectée, rétablie) dont la taille évolue en fonction du temps. On aboutit ainsi à un vecteur d'état qui décrit l'état en cours de la population et à un système d'équations différentielles (ODE en anglais) qui en définit l'évolution :



$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t)I(t) \\ \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \end{cases}$$

La figure ci-contre<sup>72</sup> montre une évolution typique de type exponentiel, caractérisée par le fameux « R effectif »<sup>73</sup> qui traduit la vitesse de croissance ou

de décroissance de cette population<sup>74</sup>. Si on veut être plus fin, on peut aussi considérer, par exemple, des sensibilités différentes selon les tranches d'âge ou d'autres paramètres. On peut aussi avoir des équations aux dérivées partielles (PDE en anglais) lorsque le compartimentage est représenté par un paramètre continu.

On notera la présence de paramètres (les lettres grecques) dans le modèle SIR ceux-ci doivent être *calibrés* pour refléter une dynamique épidémique donnée. Ceci se fait par *apprentissage* (on dit aussi *estimation*, ou *assimilation de données*) à partir de données déjà disponibles sur la pandémie en cours ou une autre, jugée similaire. On peut aussi à l'envi complexifier le modèle de base pour prendre en compte d'autres variables : géographie (répartition des cas selon le territoire), mobilités, porosité de frontières, durée d'infectiosité, etc.

Sur ce même principe, des **modèles d'agents** sont également utilisés pour décrire l'évolution d'une population à travers l'évolution de chaque agent/individu et de ses interactions avec les autres. Chaque agent possède un état (par exemple, son état sanitaire : sain, exposé, Infecté, rétabli, augmenté d'indications de position, âge, etc.). Sur la base de leur état, les agents sont susceptibles d'interagir et de modifier leur état. Ces agents se comportent donc comme des **automates**, machines dont l'état évolue au cours du temps sous l'effet d'interactions avec d'autres automates. Cette modélisation nécessite des calculs sur supercalculateur, car elle

71 Source <https://images.math.cnrs.fr/Modelisation-d-une-epidemie-partie-2.html>. On peut complexifier le modèle SIR (Susceptible, Infectious, or Recovered).en prenant des hypothèses complémentaires par exposition (SEIR), tranches d'âge, géographie, comportements, etc.

72 Dans cette figure, « retirées » est la traduction anglaise de « removed », et correspond à « rétablies ».

73 Utilisé en particulier par l'outil *TousAntiCovid*.

74 Attention, il ne faut pas confondre le caractère exponentiel de la diffusion d'une épidémie avec sa vitesse de propagation : à contamination identique (chaque cas infecte x personnes), il faut rajouter le temps (contaminer x personnes en trois jours ou trois ans ne mène pas à la même épidémie).

exige de nombreuses simulations aléatoires pour obtenir un résultat moyen et une mesure des variations.

La modélisation épidémiologique traditionnelle est aussi basée sur un concept d'homogénéité des individus et de leurs relations sociales, et sur l'hypothèse que tout individu est connecté à tous les autres. En pratique, au-delà de la simple diffusion spatiale qui limite ce réseau de relations, on observe aussi une gamme complexe de comportements, depuis les personnes isolées jusqu'aux « super-diffuseurs », dont le nombre de contacts est très élevé. Il existe des outils mathématiques (théorie des graphes, réseaux invariants d'échelle...) qui permettent d'aboutir à des modélisations plus fines et plus réalistes, mais dont les paramètres sont complexes à évaluer sur le terrain. Finalement, la question du réalisme du comportement des agents et de la représentativité du modèle reste entière.

Les deux techniques de modélisation sous-jacentes (ODE/PDE et agents) sont très largement utilisées dans d'autres domaines, en particulier des secteurs industriels tels que : aéronautique, automobile, etc., dont nous proposons de tirer inspiration.

### Épidémie, anticipations et théorie des jeux : les Mean Field Games (MFG)»

La grande faiblesse des modèles SIR est de formaliser identiquement le comportement des humains, des animaux ou des plantes, alors que les humains ont la capacité, en étant informés, de modifier leurs comportements en fonction de la situation réelle, perçue ou à venir<sup>75</sup>.

La prise en compte du facteur humain amène à d'autres outils mathématiques plus puissants relevant de la théorie de la décision. La question de la décision s'impose à toutes les échelles, de l'individu au décideur :

**C**omment sortir des paradoxes : « si tout le monde est vacciné, je n'ai pas besoin de me faire vacciner », « si tout le monde est confiné, je peux sortir » ?

- en matière macroéconomique : comment arbitrer entre des dégâts sur la santé de la population et des dégâts sur l'économie ?
- au plan de l'individu : sous quelles conditions faut-il accepter le risque vaccinal ? jusqu'à quel niveau accepter les contraintes sociales imposées ?

Ceci renvoie à la critique de Lucas, déjà citée.

Elle peut être reformulée comme suit dans le contexte de la modélisation épidémiologique :

« *On peut critiquer l'hypothèse implicite des modèles épidémiologiques selon laquelle les agents ne modifient pas leur comportement en fonction des politiques sanitaires menées, en particulier dans les études d'impact des mesures sanitaires. Il faut éviter de se baser naïvement sur des statistiques passées pour prédire le comportement futur des agents, mais prendre en compte leur réaction aux changements que les autorités vont décider, et aux annonces des autorités.*

75 À la date de l'écriture de ce rapport (mi-février 2021), on a constaté de fait une divergence entre les prévisions d'explosion de l'épidémie par les épidémiologistes et la stabilisation constatée, divergence qui a sans doute contribué à ce que le gouvernement ne suive pas les recommandations de reconfinement. Nous pensons que l'analyse ci-dessous donne des éléments d'explication à cette divergence, dont l'avenir dira si elle n'aura été que temporaire.

- « Les modèles épidémiologiques, qui sont construits par étalonnage sur valeurs passées, ne permettent pas de prévoir les effets des changements et des annonces en matière de politique sanitaire.
- « Des changements structurels (confinement, contraintes) auront des effets sur le comportement des agents. On suggère donc d'utiliser plus souvent des modèles dans lesquels les agents sont modélisés dans leurs dimensions humaines essentielles d'anticipation, de réflexion et d'adaptation de leur comportement en fonction de leur environnement. Il s'agit de modèles d'équilibre. »

Dans le monde de l'économie, la critique de Lucas a mis en exergue le concept d'équilibre entre agents dans l'esprit de la théorie des jeux. Il y a plusieurs concepts d'équilibre en théorie des jeux selon les classes de jeux. Le concept d'équilibre le plus connu, et sans doute le plus fondamental, est celui introduit par John Nash en 1950<sup>76</sup> : chaque joueur définit sa stratégie en fonction de celle des autres. Le groupe est à l'équilibre si rien ne bouge (si l'action de chaque joueur est la meilleure réponse à celle des autres). Comme les joueurs jouent simultanément, les prévisions des joueurs sur ce que vont faire les autres constituent un élément essentiel de l'équilibre de Nash.

La notion de « modèle d'équilibre » en économie est sortie transformée par la critique de Lucas : chaque agent forme des anticipations rationnelles sur l'équilibre du marché (sur les prix, les quantités...) et optimise ses décisions personnelles. Le marché est à l'équilibre si les prix anticipés induisent des décisions individuelles, donc une offre et une demande, qui s'équilibrent au prix anticipés (*prices clear the market*). Un point important est que l'interaction entre les agents économiques est médiatisée par les prix, chaque agent est en fait infinitésimal et peut être regardé comme individu d'une population. Cette globalisation fait disparaître la complexité stratégique très centrale en théorie des jeux.

La théorie des jeux à champ moyen (*Mean Field Games* - MFG<sup>77</sup>) a été proposée en 2006 par J-M. Lasry et P-L. Lions<sup>78</sup>, et indépendamment par Huang, Caines, and Malhamé<sup>79</sup>, pour fournir un cadre mathématique pour des modèles d'équilibre dans des contextes de grandes populations où les agrégats statistiques qui guident les comportements ne se réduisent pas à des prix de marché.

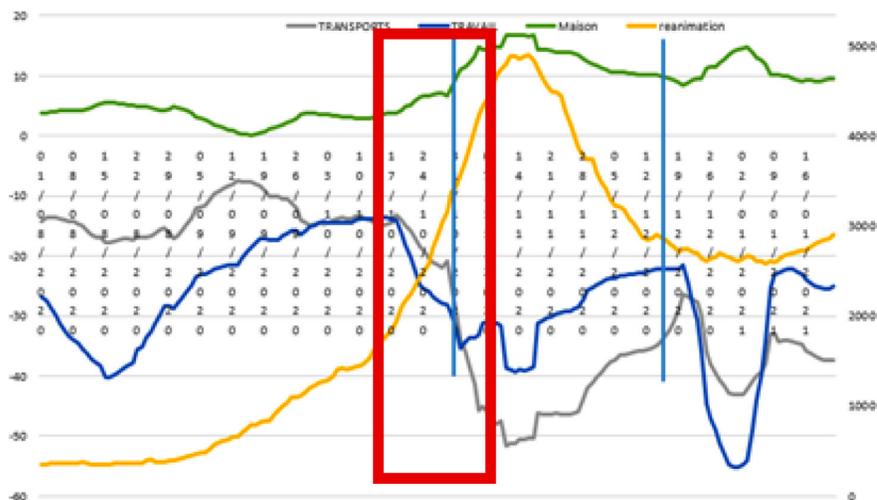
L'intérêt des modèles d'anticipation de type MFG est illustré dans le diagramme suivant.

76 [https://fr.wikipedia.org/wiki/%C3%89quilibre\\_de\\_Nash](https://fr.wikipedia.org/wiki/%C3%89quilibre_de_Nash)

77 [https://fr.wikipedia.org/wiki/Jeux\\_%C3%A0\\_champ\\_moyen](https://fr.wikipedia.org/wiki/Jeux_%C3%A0_champ_moyen)

78 Voir <https://link.springer.com/article/10.1007/s11537-007-0657-8>

79 Voir <https://link.springer.com/article/10.1007/s11424-007-9015-4>



Les anticipations de réduction de mobilité (bleu) et de télétravail (vert) démarrent 15 jours (cadre rouge) avant le confinement (barres verticales bleues) en fonction de la perception des réanimations qui augmentent (jaune). Ces anticipations et auto-restrictions sont aussi visibles dans la partie droite du schéma (fin janvier 2021).

Il ne faudrait pas conclure de ce paragraphe l'inutilité des modèles SIR : la première vague épidémique, qui a été subie AVANT la compréhension de l'influence des facteurs d'anticipation, correspond parfaitement à un modèle SIR simple.

L'utilisation des MFG pour l'évaluation anticipative de politiques sanitaires se formule comme un problème de recherche d'équilibre de type Nash, pour une grande population (éventuellement partitionnée en un petit nombre de compartiments), en réaction à des contraintes édictées par les gouvernants. Un article récent de Romuald Elie, Emma Hubert, et Gabriel Turinici<sup>80</sup> montre les écarts de prévision entre un modèle classique de type SIR et un modèle de type

MFG. Ce travail est un exemple de démarche *What-if* prenant en compte les comportements d'anticipation des populations.

De la même façon, la théorie des jeux peut s'appliquer au déploiement de la stratégie vaccinale, qui peut aller ou non dans l'intérêt (volontaire ou involontaire) des grands groupes pharmaceutiques : une vaccination lente laisse au virus le temps de muter et de déployer de nouvelles variantes, ce qui ouvre la porte à de nouveaux vaccins.

### Décision et optimisation

Dans la gestion d'une crise de type covid-19, se posent des problèmes qui conjuguent la minimisation de risques, la maximisation de survie et la logistique. La recherche opérationnelle

80 Voir <https://hal.archives-ouvertes.fr/hal-02545930v2> et tout particulièrement la Fig. 3 et son commentaire.

et l'optimisation traitent de ces questions mais dans un contexte de connaissance relativement précise des valeurs des paramètres. Or dans les grandes catastrophes, les pandémies, et les guerres, les paramètres des modèles que l'on pourrait utiliser sont partiellement ou largement inconnus. Malgré ces difficultés, il est nécessaire de prendre des décisions en quasi-temps-réel, à partir d'observations faites, elles aussi en temps réel, et des conséquences des décisions qui ont été prises.

Par exemple, en matière de confinement, on doit souvent optimiser les transferts de patients entre hôpitaux pour des raisons de surcharge (cf. l'annexe 4 qui décrit l'entretien réalisé avec IBM et qui présente, dans le cadre de la crise de la covid-19, une solution pour optimiser les transferts de patients en réanimation vers des hôpitaux moins surchargés) dans cet exemple, parmi les difficultés figurent les incertitudes concernant l'efficacité ou le moral des équipes médicales, les protocoles qu'elles suivent

**L**e point essentiel à retenir est que la multiplicité, à la fois des aspects d'une crise et des services que l'on attend des modèles, implique de savoir combiner et intégrer ces différents types d'approches. C'est précisément l'objet des plates-formes de modélisation mentionnées ci-après.

et les pertes que les transferts peuvent engendrer en matière de connaissance fine de tel ou tel patient. D'autres exemples sont l'optimisation du planning des personnels soignants dans les hôpitaux<sup>81</sup>, mais aussi l'optimisation de la vaccination (modalités de priorités, transport des vaccins, etc.). Dans ces exemples, comme dans d'autres, des mécanismes de décision décentralisée simples et rapides, fondés sur une observation et suivi en temps réel, prenant en considération les coûts et les risques, peuvent s'inspirer du secteur des enchères<sup>82</sup>.

### Analyses non numériques : fouille de textes

**L**es ordinateurs ont la capacité d'analyser aussi des textes. Cette faculté permet d'envisager de récupérer des informations quantitatives et qualitatives via les réseaux sociaux, les médias et les communications d'organisations. La **fouille de textes** (*text mining*) est une spécialisation de la fouille de données (*data mining*) et fait partie du domaine de l'intelligence artificielle. Elle consiste en un ensemble de traitements numériques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes produits par des humains et pour des humains. L'extraction de connaissances dans des textes se fait également par des techniques de *deep learning* par le biais de plongements appelés *word embedding*. Les disciplines impliquées sont donc la linguistique calculatoire, l'ingénierie des langues, l'[apprentissage machine](#), les statistiques et l'informatique<sup>83</sup>. Le suivi de la psychologie des individus s'appuie sur des outils d'**analyse de sentiment**, cf. par exemple l'étude du NIH<sup>84</sup>.

S'agissant de la crise de la covid-19, l'outil « Causality Link» présenté en annexe permet, après avoir parcouru des millions de documents, de dégager des relations de causalité entre

81 Exemple de projet <https://www.defense.gouv.fr/sante/actualites/onadap>

82 Voir par exemple <https://www.sciencedirect.com/science/article/pii/S0899825600908224>

83 [https://fr.wikipedia.org/wiki/Fouille\\_de\\_textes](https://fr.wikipedia.org/wiki/Fouille_de_textes)

84 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7591875/>

conjoncture des entreprises et chaînes de valeur à l'échelle mondiale – relations de causalité précieuses pour des analystes financiers. S'agissant de la crise du de la covid-19, ces outils ont permis de comprendre très tôt l'impact de la crise naissante en Chine et d'en anticiper, dès le printemps 2020, les effets humains et économiques à l'échelle de la planète et de chaque pays..

## MISE EN ŒUVRE NUMÉRIQUE

**D**evant cette complexité, l'usage des ordinateurs s'impose partout, que ce soit pour la diffusion des résultats et l'animation des communautés impactées (santé, économie, industrie, etc.). Les outils numériques permettent d'une part de rendre calculables les modèles décrits ci-dessus (collecte des données, leur traitement, le fonctionnement des systèmes d'analyse et prévision), et d'autre part l'utilisation et la diffusion des résultats au profit de l'organisation et de l'animation des communautés impactées (santé, économie, industrie, etc.).

On décrit ci-après quelques particularités de ces outils aujourd'hui disponibles et utilisés par les grands industriels, mais non répandus dans le monde gouvernemental.

### *Jumeaux numériques et plates-formes de modélisation*

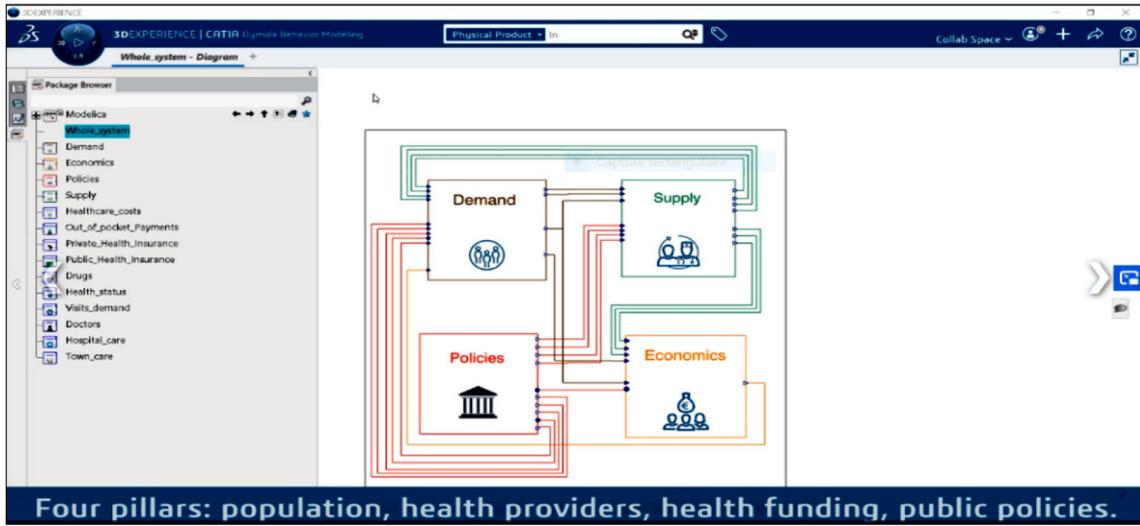
**A**ttention, le terme de *plate-forme* est devenu très surchargé de nos jours, puisqu'on parle partout de *plateformisation des services*, en écho avec « l'ubérisation de la société ». Dans le contexte du présent rapport traitant de la gestion de crises sanitaires, certaines caractéristiques des plates-formes ci-dessus restent présentes :

- la plate-forme est principalement une infrastructure d'accueil, d'intégration et de pilotage de services de modélisation et de traitements de données ;
- la plate-forme a vocation à être ouverte à des contributeurs externes, créant ainsi un écosystème.

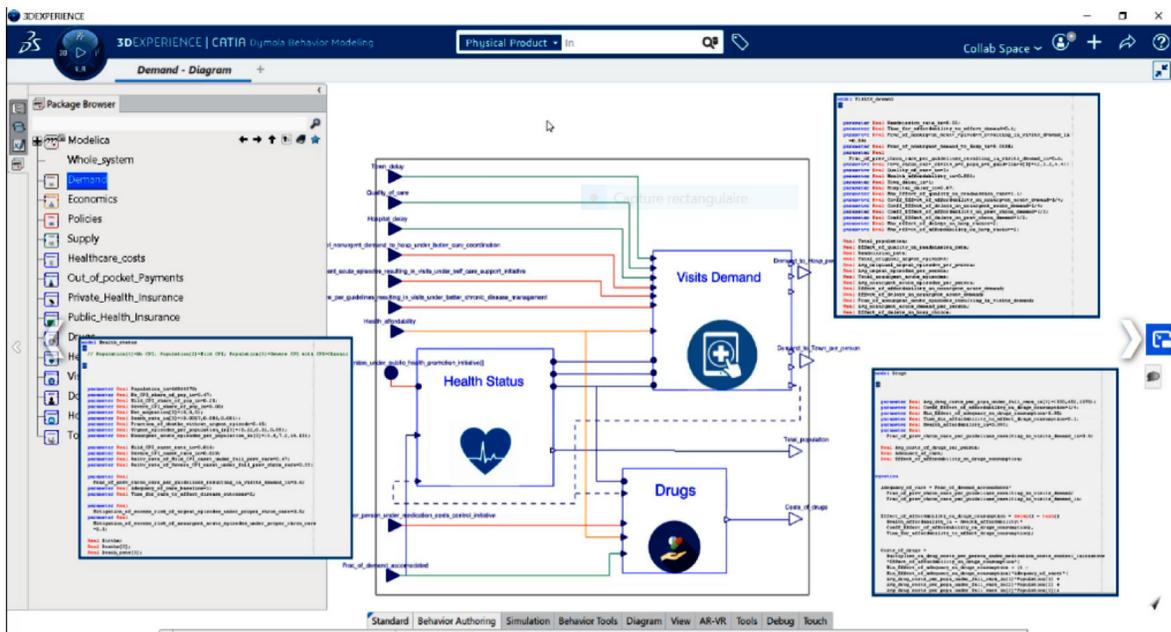
Dans notre cas, un facteur nouveau et essentiel apparaît : la plate-forme a pour colonne vertébrale une infrastructure mathématique donnant un sens précis au terme « intégration ». On développe ci-après ce point central.

**(Pour plus de détails sur ce qui suit, se reporter à l'annexe 2 sur l'entretien avec Dassault-Systèmes, dont les exemples sont extraits.)**

La figure ci-après montre la vue globale d'un jumeau numérique du système national de santé français tel que développé sur la plate-forme, et indiquant les entrées/sorties de chaque élément. Cet exemple illustre la combinaison de différents paradigmes de modélisation.



La seconde figure zoome sur la boîte « Demand ». Elle montre une vue hiérarchique correspondant à un niveau raffiné de modélisation.



Des boîtes de texte montrent un exemple du formalisme qui sert de colonne vertébrale à cette modélisation : le langage [Modelica](#), qui permet de représenter les évolutions au cours du temps de systèmes régis par des dynamiques à temps continu (équations différentielles) ou discret (agents).

Ce type de modélisation peut être associé à des techniques d'assimilation de données ou d'apprentissage, pour calibrer les modèles sur des données lorsque c'est nécessaire. Cette approche système est très générale, elle a été utilisée en biologie, pour la modélisation du cancer, pour les dispositifs médicaux, pour le génie en bioprocédé, l'économie...

L'outil offre la possibilité de définir des palettes de modèles prédéfinis et de les réutiliser. Une interface graphique rend l'outil rapidement utilisable par un non-technicien du domaine. Un code de simulation est ensuite généré, qui permet de jouer les scénarios et d'obtenir les projections *What-if*. Associé à un mécanisme flexible de gestion et d'accès à des modèles ou données (le « PLM »), ce type de plate-forme facilite le travail horizontal cohérent, en équipes distribuées.

Un concept proche est celui des plates-formes dites MDAO (*multidisciplinary design, analysis, and optimization*) destinées à explorer les grands compromis coût/service rendu/consommation énergétique en phase très amont de la conception de systèmes futurs. La NASA promeut ce concept qu'elle a contribué à développer (<https://rdcu.be/bpbhm>) et Airbus s'est doté d'une plate-forme de ce type, toujours en cours d'enrichissement (<https://hal.archives-ouvertes.fr/hal-02335530>).

Enfin, la plate-forme SDP (<https://www.thalesgroup.com/en/sdp-security-digital-platform>) de Thales, présentée en annexe 1, est un exemple de plate-forme dédiée à la gestion de la sécurité, tournée vers le traitement de données et l'analyse de situation.

### **Aide à la décision : algorithmes et outils pour les « what-if ? »**

La capacité de pouvoir évaluer des scénarii, en posant des hypothèses d'actions et d'étudier leurs impacts, s'avère un outil très important pour le pilotage de systèmes complexes, ainsi qu'en période de gestion de crise, caractérisée par des délais très courts dans lesquels il est important de prendre les décisions.

On désigne par outil « *what-if* » (que se passe-t-il si... ?) un tel outil de simulation. Il n'existe pas d'algorithme spécifique pour construire des scénarii *what-if*, sauf de préciser qu'un utilisateur est dans la boucle de décision et génère les scénarii de façon itérative pour explorer les possibilités. Toute méthode permettant de modéliser le contexte de la décision que l'on veut évaluer préalablement à son déploiement et d'en simuler les conséquences est bonne à prendre. Pour constituer une assistance efficace, cette méthode doit posséder les caractéristiques suivantes :

- la précision et la qualité de la modélisation doivent être satisfaisantes et de préférence, documentées ;
- l'entrée, c'est-à-dire la spécification de la décision, doit être formulée en des termes compréhensibles par l'utilisateur ; l'annexe 4 rapportant l'entretien avec IBM sur le thème de la logistique montre des exemples ; *a contrario*, l'obligation de coder une décision dans un jeu de paramètres sans signification pour le décideur constituerait un handicap ;
- les résultats doivent être rendus sous une forme qui parle au décideur ;
- les résultats doivent être produits dans un délai court permettant d'appuyer la décision ;
- si c'est possible, plusieurs modèles et algorithmes différents peuvent être utilisés, permettant ainsi de voir les convergences/divergences éventuelles, ce qui est un moyen heuristique minimum d'évaluer le degré de confiance ;
- l'interface doit être capable de jouer un scénario ou des milliers.

Ces caractéristiques sont précisément favorisées par le recours à une plate-forme de modélisation de type « jumeau numérique ».

## LES FORCES INDUSTRIELLES EN EUROPE

**D**eux grands industriels européens du numérique sont fortement présents dans l'offre de plates-formes de modélisation:

Dassault-Systèmes avec la plate-forme 3D EXPERIENCE (<https://www.3ds.com/fr/3dexperience>) qui unifie quatre composantes majeures : modélisation 3D et système, simulation multiphysique et multi-échelle, big data & IA et, quatrième, environnement de collaboration

Siemens avec l'outil [SIMCENTER](#) dédié à la simulation numérique et multiphysique, à côté d'outils de simulation plus spécialisés, le tout renforcé par des outils de gestion de projets de conception (PLM).

Un grand industriel du traitement de l'information, Thales, est présent sur l'analyse de situation avec sa plate-forme [SDP](#).

Enfin, des grands industriels systémiers européens ont développé, pour leur propre usage, des plates-formes pour guider les grands choix lors des phases initiales de conception des systèmes complexes (nous avons mentionné le cas Airbus, mais EDF et bien d'autres ont cette pratique). L'industrie européenne est donc riche et particulièrement solide sur ces sujets.



## Chapitre IV

### LEÇONS DES ENTRETIENS RÉALISÉS : CE QU'IL EST POSSIBLE DE FAIRE

**N**ous résumons ici les leçons que l'on peut tirer des entretiens rassemblés en annexe. Des sections précédentes, nous pouvons tirer les leçons suivantes, pour la gestion de crises telles que celle de la covid, et pour l'outillage à mettre en place. Si l'on veut résumer ces leçons en quelques mots :

**U**ne sophistication accrue des modèles n'est pas, en soi, gage de prévisions plus « certaines » de refléter la réalité. En revanche, l'intégration de modèles différents et de leurs données associées permet de rendre compte des interactions entre facettes différentes de la crise. On évite ainsi les « grosses » erreurs résultant de l'ignorance de ces interactions.

### MODÉLISATION ET DONNÉES EN SONT LA COLONNE VERTÉBRALE

#### *Disposer de données*

**G**âce au rôle central de la modélisation épidémiologique, reconnu par tous (depuis les pouvoirs publics jusqu'à l'ensemble de la population), l'importance de la modélisation en général et son caractère d'instrument essentiel pour la gestion de crises telles que celle de la covid-19, sont acquis. Pour être pertinents et correctement prédictifs, les modèles ont besoin de données : pour l'épidémiologie, l'utilisation de modèles de type « semi-physiques<sup>85</sup> » requiert en

---

85 On dit aussi « modèle physique simplifié » il s'agit de modèles dont la structure reflète, avec des simplifications, des connaissances physiques sur le phénomène décrit et comporte divers paramètres réglables. L'ajustement de ces paramètres sur des données collectées (parfois appelé « calibration ») permet de mieux reproduire les phénomènes observés. La calibration, combinée avec le caractère semi-physique de la modélisation, fournissent en général un bon pouvoir prédictif, fréquemment supérieur à ce que peuvent faire les modélisations purement « boîte noire » (n'impliquant aucune information de nature physique) utilisées par exemple dans l'apprentissage profond ceci vaut tout particulièrement lorsque les données sont en nombre insuffisant pour entraîner un réseau de neurones profond et qu'il n'existe pas de moyen d'entraîner au moins en partie ce réseau sur des applications comparables et bien dotées en données.

permanence une calibration (ajustement des paramètres) sur des données mesurées pour améliorer leur pouvoir prédictif. Les données statistiques globales sur l'évolution de l'épidémie sont actuellement récoltées et publiées en nombre, voir l'annexe technique et bibliographique — 3. Sur les données, qui donne quelques exemples.

En revanche, mener des efforts de modélisation plus globaux, incluant d'autres aspects, demande la disponibilité d'autres données liées au système de santé et aux usages connexes pouvant influencer sur la pandémie. Ainsi, l'Inserm et Orange ont établi au printemps 2020 un partenariat pour publier des données agrégées de mobilité grâce aux mesures réalisées sur le réseau mobile Orange<sup>86</sup> il ne semble pas que ce service ait connu un réel développement.

Plus important, on a la chance en France d'avoir une seule organisation d'assurance maladie (la Sécurité sociale) qui a toutes les données de tous les patients. Donc on pourrait tout explorer, en principe, modulo le respect de la protection de la vie privée — et justement, le service données-mobilités cité plus haut respectait les exigences de l'anonymisation<sup>87</sup> —, on pourrait s'en inspirer. Tous les actes médicaux sont archivés par la Sécurité sociale (fichier Sniiram<sup>88</sup>). Cet avantage est unique comparé aux autres pays. Malheureusement, ces données de la Sécurité sociale ne sont pas disponibles à des fins de modélisation, même si le [health-data hub](#) se positionne comme un portail d'accès fédéré à divers entrepôts de données.

### *Croiser les modélisations des divers aspects de la crise*

Comme nous l'avons vu, la gestion de la crise due à la covid-19 prend en compte d'autres facteurs que la seule dynamique de l'épidémie :

- pour la lutte contre la pandémie, les aspects liés au support industriel et à la logistique sont importants (mise au point, disponibilité et distribution des masques, tests et, aujourd'hui, vaccins) ;
- l'impact des mesures (confinement, couvre-feu, restrictions de déplacement) sur le fonctionnement de l'économie s'est révélé être un paramètre essentiel. C'est en particulier la raison principale du maintien de l'ouverture des écoles par le gouvernement français. Il en va de même pour toutes les mesures de type confinement et couvre-feu ;
- le degré d'acceptabilité des diverses mesures par la société, variable selon les pays ou régions du monde, est une contrainte importante. C'est sans doute cet aspect qui prédomine dans les différences entre les pays et régions quant à la maîtrise de la pandémie.

### *Croiser les informations pour préparer les décisions*

Différents corps de l'État (ministères et administrations) doivent, conjointement, élaborer leurs décisions. Il ne nous appartient pas de formuler des recommandations sur l'organisation (au-delà du lapidaire « casser les silos »), mais nous allons en tirer les conséquences quant à

86 Voir <https://presse.inserm.fr/les-statistiques-issues-du-reseau-de-telephonies-mobiles-au-service-de-la-lutte-contre-la-pandemie-de-covid-19/38831/>

87 Prenons cette affirmation avec quelques pincettes : on voit régulièrement ces procédures d'anonymisation cassées par des analyses statistiques poussés autant qu'ingénieuses, mais sans doute rarement applicables en routine.

88 Voir <https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/sniiram/finalites-du-sniiram.php>

l'outillage dont une nouvelle organisation aurait besoin.

## PLATES-FORMES, APPROCHE COCKPIT, JUMEAU NUMÉRIQUE, DONNÉES : ÉTAT DE L'ART

**A**u total, la demande et les besoins issus des secteurs industriels indiqués plus haut ont abouti au développement d'une infrastructure mathématique et informatique très flexible en matière de modélisation :

- modélisation multi-physique (aérodynamique et vibrations, électricité et thermique...)
- modélisation conjointe de la physique du système et du logiciel qui le contrôle ;
- modélisation de type système de systèmes, par exemple pour les systèmes de transport (mobilité par système de véhicules autonomes).

Le développement de ces techniques de modélisation et de leurs combinaisons a fait émerger des solutions de modélisation ouvertes et flexibles.

Les plates-formes en question offrent différents services repris dans les paragraphes suivants.

### *Une interface ergonomique pour utilisateur non nécessairement spécialiste*

**E**lle peut être déployée, selon les besoins et configurations, soit sur un poste personnel, soit sur une salle de commande (telle que celle dont une photo est donnée dans l'entretien Thales, voir l'annexe 1). Le point essentiel est que les résultats de la modélisation sont retournés à l'utilisateur *sous le format qui lui parle le mieux, selon la nature de l'objet modélisé*. Par exemple, s'il s'agit d'un système mécanique (voiture, avion...) pour lequel on souhaite étudier l'aérodynamique, les flux d'air sont rendus et visualisés « au contact de la forme ». Dans le cas de la covid-19, des simulations de salle ont été réalisées, et la circulation des flux d'air de la salle a été rendue sur une visualisation en 3D de la salle (pour l'étude prédictive de la transmission du virus). S'il s'agit d'un système de logistique (optimisation de la relocalisation de patients en réanimation), les résultats sont retournés sur une carte avec surimpression des transports optimaux (voir l'entretien conduit avec IBM, annexe annexe 4).

De plus, la même crise peut être inspectée sous des aspects différents (épidémiologie, logistique, société), et, pour ce faire, les résultats de la même modélisation peuvent être renvoyés selon des styles de visualisation différents, adaptés aux différents aspects. Chaque acteur voit ce qui le concerne, dans l'idiome qu'il comprend, mais tous les acteurs voient bien le même objet.

Cette approche est parfois qualifiée d'**approche cockpit**, en référence au poste de pilotage d'un avion.

### *Une capacité puissante de captation des données*

**C**es plates-formes sont dotées de capacités à récolter les données et à les héberger. Typiquement, les données intéressantes peuvent être de trois catégories (voir les entretiens avec

Thales en annexe 1 et l'entretien avec Causality Link en annexe 3) :

- obtenues par saisie manuelle, quand aucun autre moyen n'existe et que les données en question justifient le coût d'une telle démarche ; ces plates-formes sont donc dotées d'interfaces facilitant une telle saisie. La collecte collaborative (crowdsourcing) peut être incluse dans cette catégorie ;
- récoltées automatiquement à partir de bases de données ouvertes (telles que les données publiées par le site [www.santepubliquefrance.fr](http://www.santepubliquefrance.fr) s'agissant de la pandémie). Cette récolte présente de sérieuses difficultés en raison de leur nature non standardisée (dès qu'on collecte des données de différentes sources ouvertes), voire non structurée (documents, planches...) ;
- provenant de flux ouverts accessibles (réseaux sociaux), ou de flux accessibles commercialement. Sur ce dernier point, on peut citer comme exemples : la plate-forme OpenHealth de suivi de la consommation de médicaments (à l'échelle française et mondiale), et celle de Causality Link pour les données économiques et causalités.

### **Des ressources importantes et diversifiées en termes de modélisation**

Ces plates-formes intègrent nombre de composants algorithmiques permettant modélisation, analyse, et aide à la décision, dans des domaines variés, couvrant entre autres les volets suivants :

**U**n point clé dans la gestion de crises inédites est que l'on découvre au cours de la crise quels sont les facteurs essentiels qui interviennent et sur lesquels les autorités peuvent interagir. Il est donc illusoire d'espérer être pleinement préparé et outillé.

- modélisation multi-physique pour les systèmes combinant mécanique, électricité, thermique, fluide, magnétisme, etc.
- épidémiologie, avec une palette de modèles : modèles de type SIR à base d'ODE (équations différentielles ordinaires), le plus souvent compartimentés, mais aussi modèles dits « multi-agents », où la propagation du virus est modé-

lisée au niveau des individus, de manière probabilisée, ce qui permet de différencier plus finement les individus selon leur comportement ;

- logistique et recherche opérationnelle, avec des méthodes de modélisation et des procédures d'optimisation ;
- et jusqu'à la modélisation même de l'infrastructure de calcul embarquée, le cas échéant (important pour une automobile ou un avion, mais sans objet pour la crise de la covid-19).

Ces différentes facettes sont intégrées pour être combinées. Elles peuvent accéder à des données communes. Dans certaines plates-formes même, ces diverses facettes peuvent être développées au-dessus d'une base de modélisation commune, permettant ainsi leur synchronisation (épidémiologie et logistique peuvent ainsi se parler, par exemple).

On voit ainsi émerger la notion de **jumeau numérique**<sup>89,90</sup>, (*digital twin* en anglais), deve-

89 Voir par exemple [https://www.maplesoft.com/ns/manufacturing/industry-4-0-power-of-the-digital-twin.aspx?p=TC-5626&gclid=EAIaIQobChMlxo2l5ven7glV1pTVCh1RbADWEAMYASAAEglo6\\_D\\_BwE](https://www.maplesoft.com/ns/manufacturing/industry-4-0-power-of-the-digital-twin.aspx?p=TC-5626&gclid=EAIaIQobChMlxo2l5ven7glV1pTVCh1RbADWEAMYASAAEglo6_D_BwE)

90 [https://fr.wikipedia.org/wiki/Jumeau\\_num%C3%A9rique](https://fr.wikipedia.org/wiki/Jumeau_num%C3%A9rique)

nue centrale aujourd'hui chez les grands industriels systémiers. Les jumeaux numériques sont de plus en plus prévalents dans le secteur aéronautique (avec le concept d'iron bird, l'avion existant sous forme de modèle numérique avant que d'être effectivement construit) et ils deviennent absolument incontournables lorsqu'il s'agit d'explorer des architectures nouvelles d'avion (propulsion électrique et/ou hydrogène, propulsion distribuée sur de nombreux petits moteurs, aile-avion), pour lesquelles les extrapolations à partir des réalisations passées sont impossibles.

Or, justement, les crises telles que celle de la covid-19, sont inédites, et, à ce titre, justiciables du recours aux jumeaux numériques.

## CAPITALISER L'EXPÉRIENCE ET LES MEILLEURES PRATIQUES

Comme nous l'avons souligné, la qualité de la gestion de crise en Europe a été d'une grande diversité (la France se situant en position médiane). Le problème n'est pas de distribuer des bons et mauvais points, mais de retenir les éléments positifs pour la prochaine crise :

- anticipation et préparation de la crise à l'échelon national ;
- rapidité de la réactivité ;
- formation des gestionnaires et personnels de crise ;
- partage de l'information.

Tout ceci nécessite des outils et plates-formes informatiques configurables rapidement, tels que décrits dans ce rapport. Nous sommes enclins à penser que l'usage de telles plates-formes conduirait naturellement à la capitalisation d'expérience.

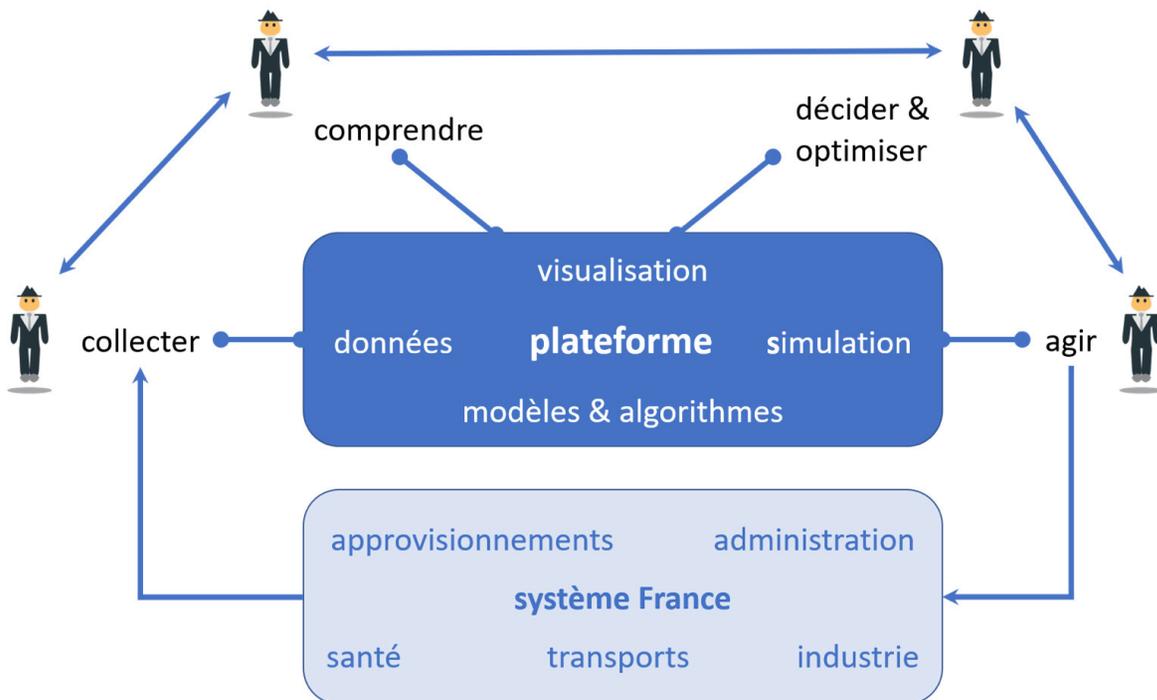


## CONCLUSION

### POUR UNE PLATE-FORME DE MODÉLISATION POUR LES CRISES SANITAIRES

I s'agit de mettre en place une nouvelle façon d'analyser, de décider, et de communiquer, basée sur des outils partagés via une plate-forme.

De façon générale, une plate-forme a une architecture et un usage correspondant au schéma suivant :



La plate-forme se comporte comme une collection d'outils, activables simplement et efficacement par des équipes situées en des lieux variés. Une caractéristique très importante est la souplesse d'évolution, la modularité et la capacité à refléter des granularités multiples (échelons administratifs, entreprises spécialisées, organisations de soins, logistique, etc.). Une

plate-forme n'est pas nécessairement monolithique : elle peut être activée via des interfaces (APIs) informatiques rendant l'implémentation physique transparente à l'utilisateur.

Comme le montrent les notes des entretiens rassemblées dans les annexes, la France est particulièrement bien dotée en matière d'offre correspondant à ce concept.

## COMBATTRE L'IMAGE DE « DÉCISIONS-GIROUETTE »

**D**ans toutes les crises, et tout particulièrement celle du de la covid-19, toutes les décisions se prennent en présence de fortes incertitudes. Les effets de ces décisions sont visibles avec un retard significatif et sont probablement elles-mêmes prises de manière plus ou moins aléatoire. Ce type de situation est très bien connu et compris dans la théorie mathématique de la décision. Dans ce contexte, la nature de la décision recommandée (dite « optimale ») est qu'elle est *mixte* (c'est-à-dire aléatoire elle-même) : dans le cas simple où la décision se fait entre les deux termes d'une alternative, on peut voir la décision comme un tirage biaisé à pile-ou-face, où c'est le biais qui est optimisé (faisant que le choix de pile sera, par exemple, plus probable que celui de face). Dans cette démarche, une légère modification des données mesurées conduit à une légère modification du paramètre caractérisant le biais. Le mode de décision n'est donc pas celui d'une girouette.

Pour illustrer ce mécanisme par l'alternative confiner/ne pas confiner, il faudrait, *stricto sensu*, donner ce biais à chacun des individus, chacun tirant ensuite au hasard son comportement à l'aide du pile-ou-face ainsi biaisé. De la sorte, la population se comporterait statistiquement selon la politique optimale. Évidemment, aucune société n'implémenterait une telle politique. En pratique, c'est le politique qui transforme cette décision entachée d'incertitude, en une décision ferme, affichée comme étant « la meilleure » (en choisissant, de pile ou de face, le plus probable selon la stratégie optimale). Il peut difficilement en aller autrement. Du coup, un changement léger dans les données recueillies peut se traduire en un revirement brutal de la politique suivie, ce qui est interprété fréquemment comme une inconsistance par le public (et les médias). Une perte de confiance et donc d'efficacité peut en résulter. Y a-t-il un moyen d'échapper à ce schéma délétère ? Expliquer une décision par l'exposé de scénarios « *what-if* » (voici les conséquences probables de telle décision) alternatifs limite l'écueil ci-dessus. Mettre sur la table les termes de l'alternative et en exposer les conséquences (grâce à la modélisation) réduit les risques d'apparaître comme une girouette en cas de changements opportunistes mais justifiés, de politique.

**C**ette attitude de « mettre les éléments sur la table » par l'exposé de scénarios *What-if* est le premier élément important. Le recours à une plate-forme le permet naturellement.

## PRENDRE EN COMPTE L'ENSEMBLE DES ÉLÉMENTS POUR FAIRE ÉMERGER LA MEILLEURE STRATÉGIE

Les décisions d'action en cas de crise résultent de compromis entre des conséquences contradictoires (on l'a bien vu avec la covid). Les éléments de ces compromis sont souvent remontés par des acteurs différents de la cellule de décision si l'examen de ces éléments se fait sous forme trop synthétique (les « notes de synthèse »)

les facteurs aboutissant à une bonne pondération peuvent avoir été gommés. Rendre en permanence ajustable le degré de détail ou au contraire d'abstraction de ces facteurs rendra plus pertinente la décision. C'est un second service que rend naturellement le recours à une plate-forme.

Le but d'une plate-forme n'est pas seulement d'optimiser les décisions, il est aussi et surtout d'éviter les erreurs.

## DONNER LES ÉLÉMENTS EXPLICATIFS DE LA DÉCISION

Ce besoin a déjà été clairement compris par les autorités gouvernementales, on le voit dans les conférences gouvernementales rapportant l'état de la pandémie et les décisions. Les facilités de support et d'affichage des résultats offerts par une plate-forme telle que résumées chapitre 4 paragraphe : 5.2 : *Capitaliser l'expérience et les meilleures pratiques* fourniraient de bonnes illustrations pour argumenter les décisions.

## AVEC QUELLE ORGANISATION ?

Comme nous l'avons déjà souligné, notre groupe de travail ne se sent pas qualifié pour formuler des propositions en matière d'organisation. Il convient donc d'examiner si le recours à un outillage tel que les plates-formes et jumeaux numériques que nous avons présentés et défendus, impose des contraintes sur les modes d'organisation. **Notre opinion est qu'il n'en est rien.** Examinons pourquoi.

### *Combiner et synchroniser des paradigmes de modélisation différents*

Les plates-formes sont des infrastructures *ouvertes* sur différents paradigmes de modélisation : orientés modèles, orientés données, mélange des deux. Elles n'imposent pas de paradigme en particulier.

### *Exposer à chaque intervenant une vue adaptée sur modèles et données*

Chaque acteur se voit présenter la vue qui le concerne sur le système modélisé. Ce mécanisme de vues est paramétrable et non pas figé il est conçu à la manière d'une gestion flexible de droits d'accès.

**D**e la sorte, la question du choix de l'organisation (centralisée, décentralisée, mélange) demeure indifférente au déploiement de l'infrastructure numérique qui l'accompagnera. Les arguments en faveur de telle ou telle organisation peuvent donc être évalués indépendamment, sur d'autres critères.

un point important à souligner, d'autant plus que l'existence de telles contraintes est un reproche qui est fait fréquemment à l'encontre de supports numériques.

## **S'APPUYER SUR LES ACQUIS INDUSTRIELS ET SUR LA RECHERCHE**

**L**a crise a fait apparaître les forces et faiblesses de l'organisation française et européenne. Parmi les forces, on peut citer :

- l'éventail considérable de compétences disponibles tant dans le matériel (fabrications diverses, capacités de tests vétérinaires...) que dans l'immatériel (numérique, services, logistique, etc.) ;
- la réactivité et adaptabilité : des chaînes de valeur ont été arrêtées, reconverties ou mises en œuvre en l'espace de quelques semaines.

Ces forces doivent être des appuis pour la reprise, et aussi pour les crises qui pourraient advenir.

Enfin, si les développements de type plate-forme sont plus particulièrement dans le giron de l'industrie, certains besoins de modélisation relèvent clairement de la recherche : là aussi, croiser les compétences et faire se rencontrer les communautés est indispensable.

## ANNEXES

Liste des annexes :

- Annexe 1 — entretien avec Thales
- Annexe 2 — entretien avec Dassault-Systemes
- Annexe 3 — entretien avec Causality Link
- Annexe 4 — entretien avec IBM
- Annexe 5 — entretien avec Jean-Michel Lasry
- Annexe 6 — entretien avec OpenHealth sur le réseau COMETE
- Annexe technique et bibliographique
  - État de la pandémie
  - Sur la modélisation épidémiologique
  - Sur les données
  - Aspects économiques, sociaux et sociétaux

## ANNEXE 1 — ENTRETIEN AVEC DES MEMBRES DE LA SOCIÉTÉ THALES

(27 Novembre 2020)

Ces notes prises Albert Benveniste et Laurent Gouzènes ont été relues par M. Makrem Dridi

### ➤ *Participants*

#### Académie des technologies

ALBERT BENVENISTE

LAURENT GOUZÈNES

GÉRARD ROUCAIROL (excusé)

#### Thales

MAKREM DRIDI

LAURENT LETELLIER (chef de produit de la SDP)

HERVÉ NOËL (architecte système, construction de solutions)

MARKO ERMAN

#### *Convention :*

*Les questions sont indiquées en italique et les réponses en roman les auteurs des questions sont indiqués par leurs initiales.*

## A. UNE PLATE-FORME POUR LA GESTION DE CRISE : SECURITY DIGITAL PLATFORM (SDP)

EXPERTISE DU GROUPE THALES ET RÉFÉRENCES

### References

URBAN SAFETY & MOBILITY	LARGE EVENTS & CRISIS MANAGEMENT	AIRPORTS	MILITARY SITES
 <p><b>MEXICO</b> The world's most sophisticated urban security system Crime rate reduced by 55% Response times cut divided by 3</p>	 <p><b>MECCA</b> Complete security, telecom and crowd management system 2 million pilgrims</p>	 <p><b>JFK/ T4 AIRPORT</b> Deployment of state-of-the-art security technology 20 million+ passengers per year</p>	 <p><b>FRENCH MoD HQ</b> High-performance security and telecom system</p>
 <p><b>DUBAI SMART CITY</b> Advanced multi-modal, multi-agency platform for transportation 2020-2021 2020-2021 2020-2021</p>	 <p><b>EUROPE (Undisclosed)</b> National crisis management COVID Crisis</p>	 <p><b>DUBAI AIRPORT</b> Ultra-modern security and communication solution 88 million+ passengers in 2017 10,000+ access control points</p>	 <p><b>NETHERLANDS MoD</b> Swap 45 systems by 1 multi-site solution 170 military sites 15 years of operations</p>
 <p><b>NICE &amp; PARIS BUSINESS CENTER</b> Collaboration city security's stakeholders Integration &amp; interoperability between +30 systems &amp; solutions</p>	 <p><b>MIDDLE EAST (Undisclosed)</b> National Crisis management centre State security, safety, welfare</p>	 <p><b>OMAN</b> New exceptional airport infrastructures 40 flights per hour with 86 check-in counters and 40 gates</p>	 <p><b>SUB MARINE BASE</b> High-performance Access control and alarm security</p>

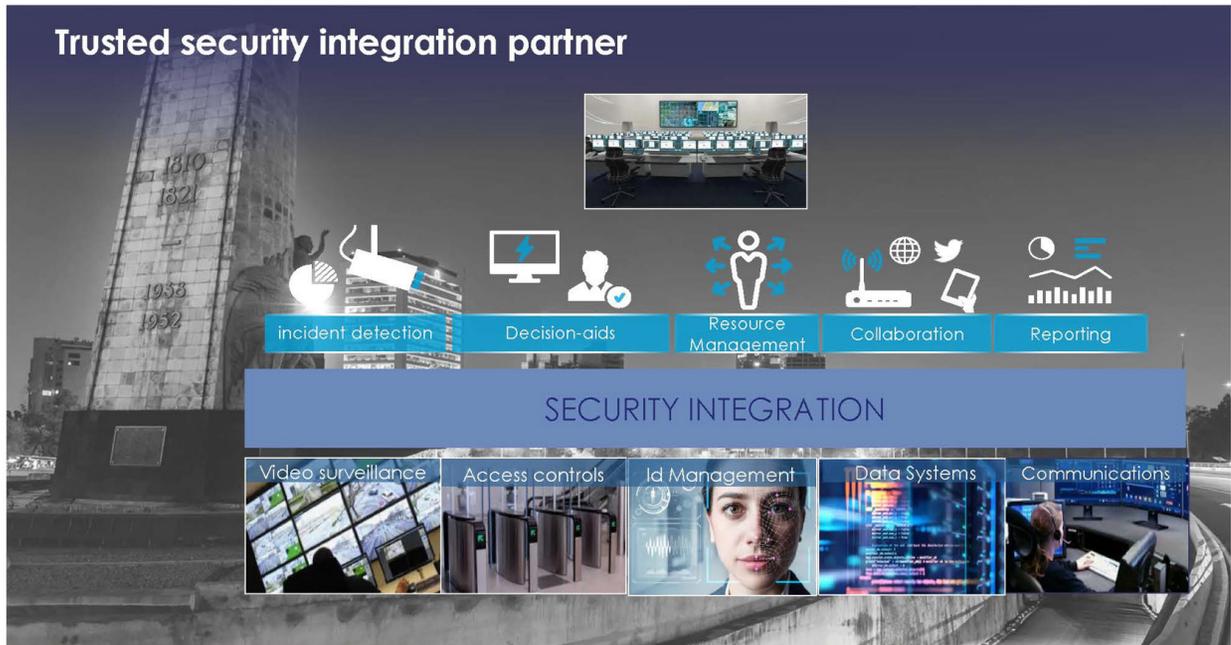
4

OPEN

THALES

Nous voyons ci-dessus les cibles originelles et potentielles de la plate-forme Thales SDP, lancée en 2018. Toutes comportent un aspect gestion de crise. Il s'agit souvent d'anticiper des grands événements tels que l'exposition universelle de Dubai. Thales sécurise le site de La Mecque avec des fonctionnalités avancées de gestion de foule et de traitement des incidents — Thales reçoit de plus en plus de demandes liées à la gestion de crise, en particulier au Moyen-Orient. Thales assure la sécurité des aéroports — y compris comment y gérer les accès COVID. Pour les sites militaires, Thales gère de gros projets aux Pays-Bas et en France. Thales a déployé quelques cas de plates-formes COVID, aux niveaux État et ville. De façon générale, Thales développe beaucoup d'IA pour la reconnaissance de formes, ce qui est un élément clé pour les activités de gestion de crise.

L'ADN de Thales est dans l'intégration de la sécurité : gérer de nombreux sous-systèmes pour intégrer, consolider et apporter une intelligence.



➤ *Sécurité pour la ville de Mexico : origine de Thales sécurité*

### Security for Mexico City Mexico

**Customer needs**

- Increase the Security and Safety for 9 million citizens

**Thales solution**

- Fully integrated Command Centres :
  - 1 central, 5 regional, 2 mobile
- Extensive security management unified suite:
  - Call taking, Incident Management, Call Aided Dispatch, Reporting, Video Management, ...
- Multi-Agency Concept of Operations

**Benefits**

- Increase in tourism: 47%
- Decrease of unemployment: 28%
- Crime rates reduced by 56% during the 2010-2016 period
- Response times: from 12 min to 4 min.

**Vast network of sensors deployed:**

- 20 000+ CCTV cameras
- 850 license plate recognition
- 9,800 emergency call points,
- 36 gunshot sensors, UAV...

Over 40,000 calls and 5,000 emergencies handled every day

➤ *Thales a ensuite évolué vers la mobilité : Dubai*

## Smart Mobility, Dubai UAE

### Customer needs

- Enhance the quality of service, safety and efficiency of the Public Transport services of Dubai

### Thales solution

- The world's first multi-Modal, multi-Agency command and control centre to manage all transport means
- Outcome: Dashboards, Incident & Crisis management, event management, prediction & modelling, Information dissemination, CCTV management, multi-layer GIS

### Benefits

- Improved incident and crisis management, event management, capacity & infrastructure planning, Security & Safety and user services



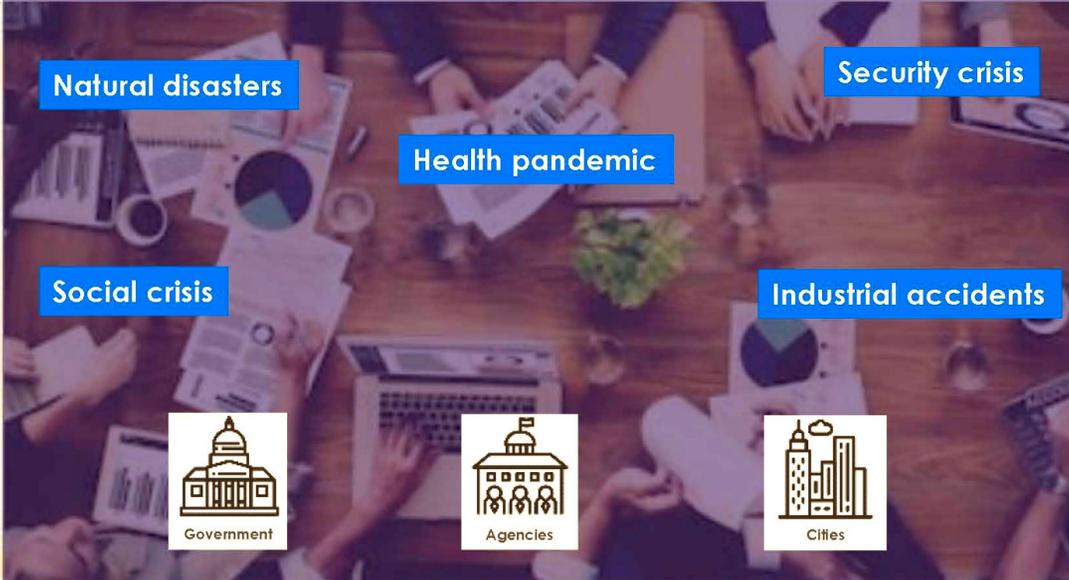
Collection of data from over 34 systems, processes and analyses approximately 75 million data records per day

7
OPEN
THALES

➤ *Thales est finalement arrivé à la notion plus générique de gestion de crise*

## Crisis Management

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales 2015. All rights reserved.





Government



Agencies



Cities

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales 2015. All rights reserved.

OPEN
THALES

## EXAMPLES OF CRISIS MANAGEMENT FUNCTIONS

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. - © Thales 2016. All rights reserved.



Large events supervision



Crowd management



Social networks monitoring



Advanced dashboards



Big data ingestion



Crowd analysis



Unified data lake



Decision-aid



Speedy roll out

12

SWP P1 - 04/2019 v2

OPEN

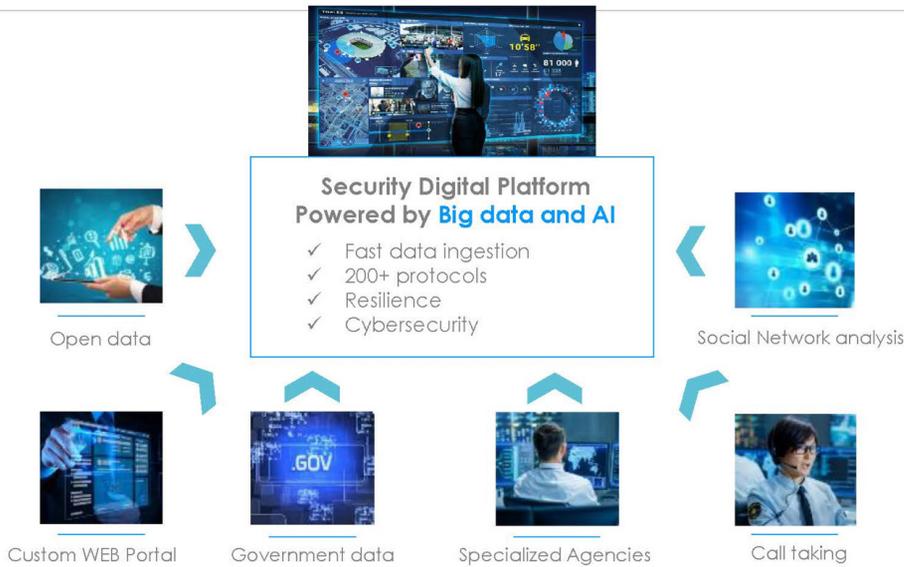
THALES

La gestion de crise a beaucoup de besoins en commun avec les cas précédents. En outre, il faut inclure de l'analyse de sources ouvertes (réseaux sociaux), de la détection de signaux faibles, des analyses de sentiment. Cela implique une capacité à récupérer toutes ces données et les mettre à disposition des clients pour qu'ils produisent eux-mêmes leurs propres tableaux de bord et leurs analyses (nature des clients de Thales aujourd'hui : principalement institutionnels, étatiques, grandes villes).

### LA PLATE-FORME SDP

## CRISIS DIGITAL PLATFORM

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. - © Thales 2020. All rights reserved.

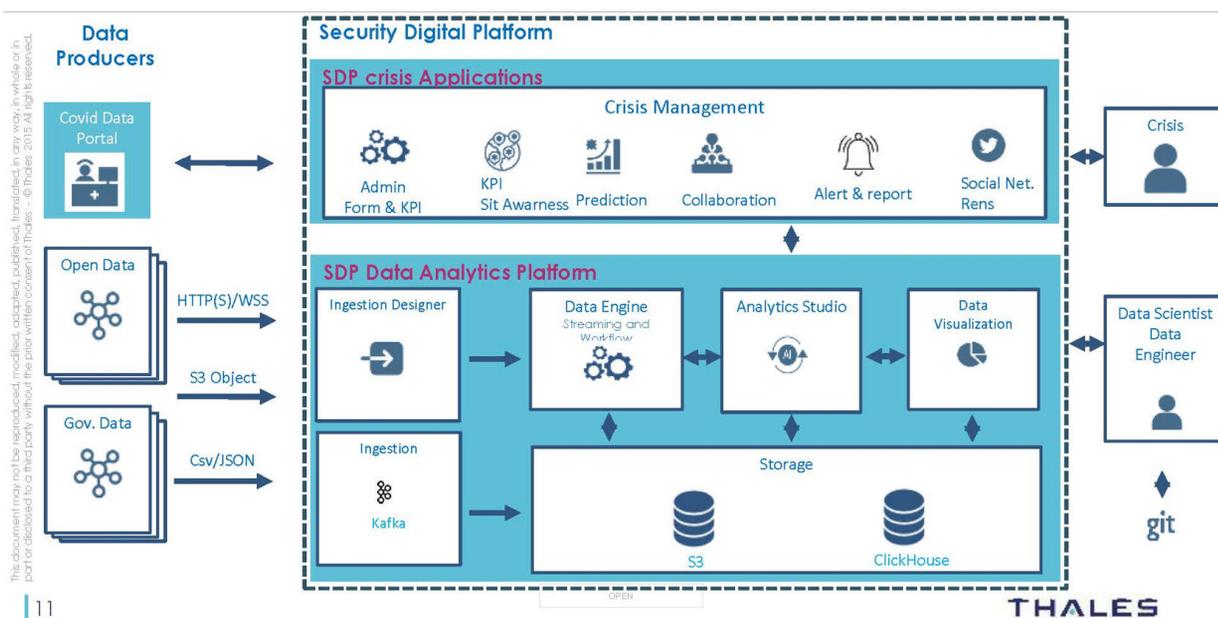


10

THALES GROUP INTERNAL

Un point clé est la capacité à récupérer les informations de sources diverses. Il faut permettre l'entrée manuelle des données quand on n'a rien d'autre. SDP offre donc des interfaces facilitant cette saisie. Elles sont adaptées à la diversité des utilisateurs, voir la section « hétérogénéité ». SDP collecte les données institutionnelles et étatiques, voir à ce sujet la section « hétérogénéité ». Les réseaux sociaux sont intéressants, ainsi que les appels d'urgence. Toutes ces informations sont automatiquement captées par le système et ingérées pour traitement. Il faut donc avoir l'agilité pour se déployer sur les divers environnements des clients.

## TECHNICAL PLATFORM



La plate-forme SDP est basée sur des standards qui permettent d'intégrer des données et des composants offrant divers services (traitements en *streaming* ou *batch*). Analytics Studio est l'outil de développement des modèles. L'agilité est fondamentale : un déploiement est effectué tous les quinze jours. Il faut pouvoir se déployer sur des plates-formes quelle que soit leur architecture. L'architecture de la plate-forme SDP est innovante et évolutive : elle utilise des micro-services et des services de stockage de données pour ne pas être limitée dans les changements d'échelle.

SDP utilise Argo pour faire des lignes de traitement en *batch* génériques, et JupyterHub pour les lignes de traitements de modélisation, qui permettent de faire collaborer des outils et composants. Un modèle commun existe pour la structuration des données, avec des contraintes imposées concernant le format et la méthode de stockage. On distingue sources brutes, résultats de traitements, métadonnées.

➤ *Vue d'une salle d'analyse*

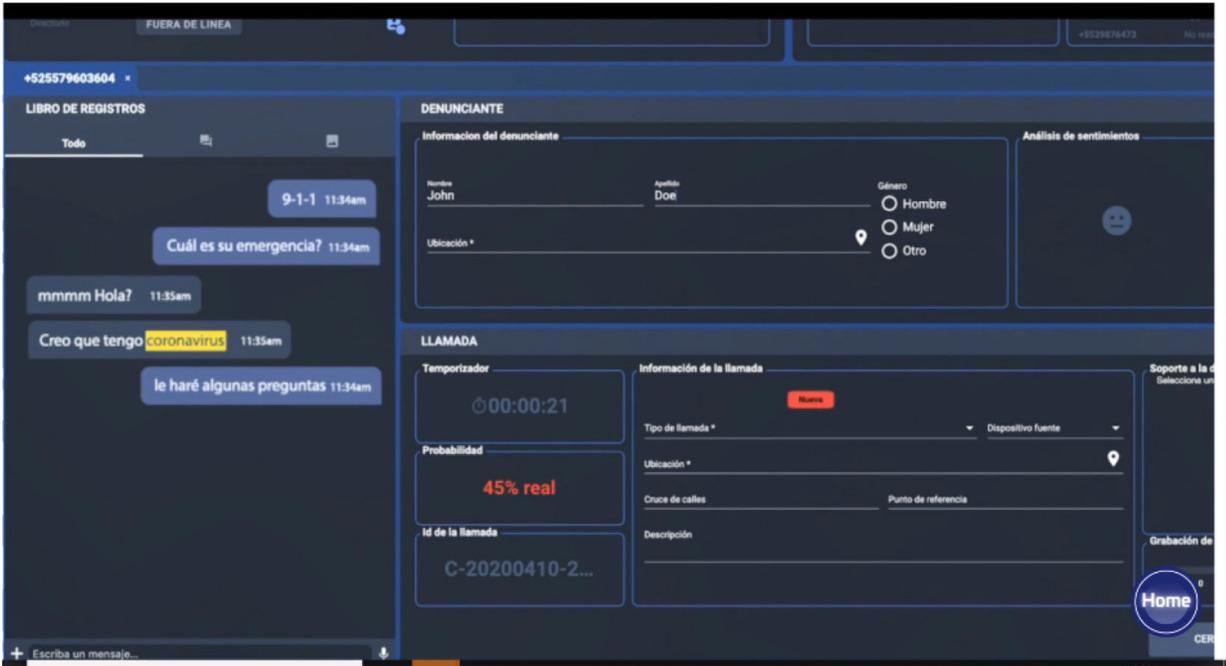


La salle ci-dessus est typiquement faite pour la gestion des incidents, problème assez décentralisé. D'autres crises (décision de type « confinement » comme dans la pandémie de la covid-19) sont des décisions beaucoup plus centralisées. Thales est souvent acteur dans la définition de ces rôles opérationnels. Ce type de plate-forme est typiquement pertinent pour assister le Conseil de défense renforcé lors du COVID — ceci existe déjà depuis longtemps dans le secteur de la défense, avec des caractéristiques proches.

Toutefois, la gestion de la crise covid est bipolaire : santé et conseil de défense, ce qui complique la chose. Mais, justement, partager le même outil au niveau horizontal permet de mieux partager l'information et la compréhension. Si une partie ne voit pas ce que voit l'autre, on a du mal à se mettre d'accord. On essaie ainsi de faire tomber les cloisons entre ces diverses administrations ou ministères.

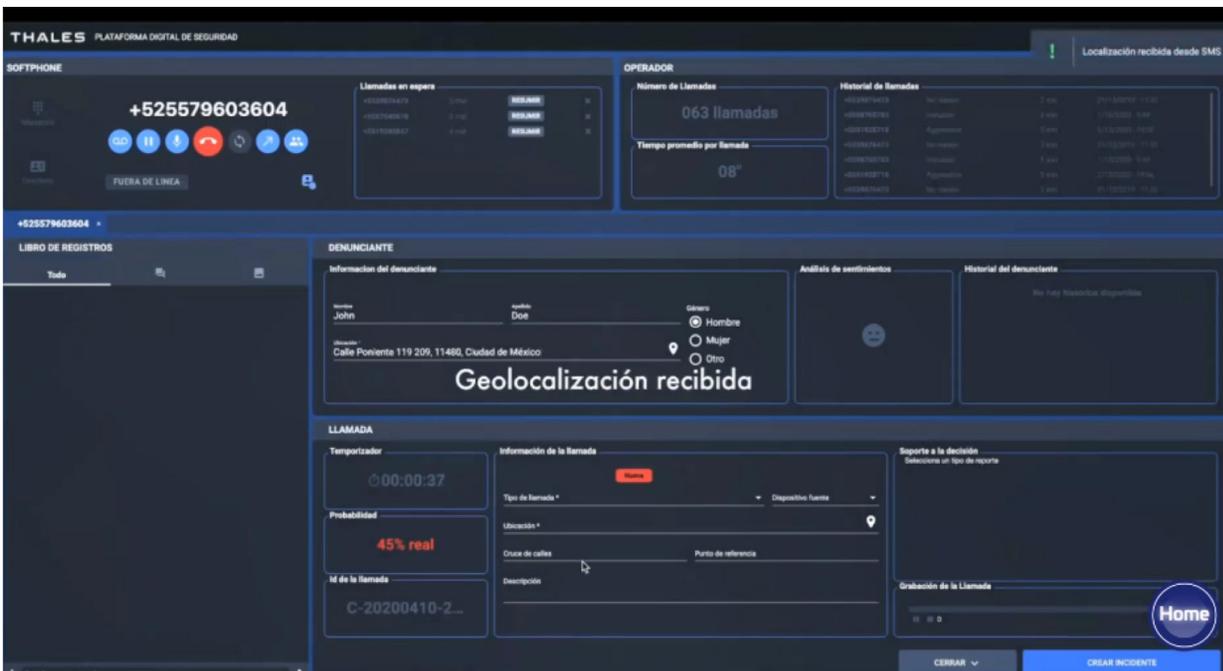
En référence à la défense, dans les opérations interarmées dans le passé, on remontait très haut dans les hiérarchies les informations, pour décider et redescendre, ce qui n'est pas efficace. Maintenant on crée des communautés d'intérêt temporaires où l'on partage informa-

tions et décisions. La numérisation du champ de bataille ouvre la voie vers cela.

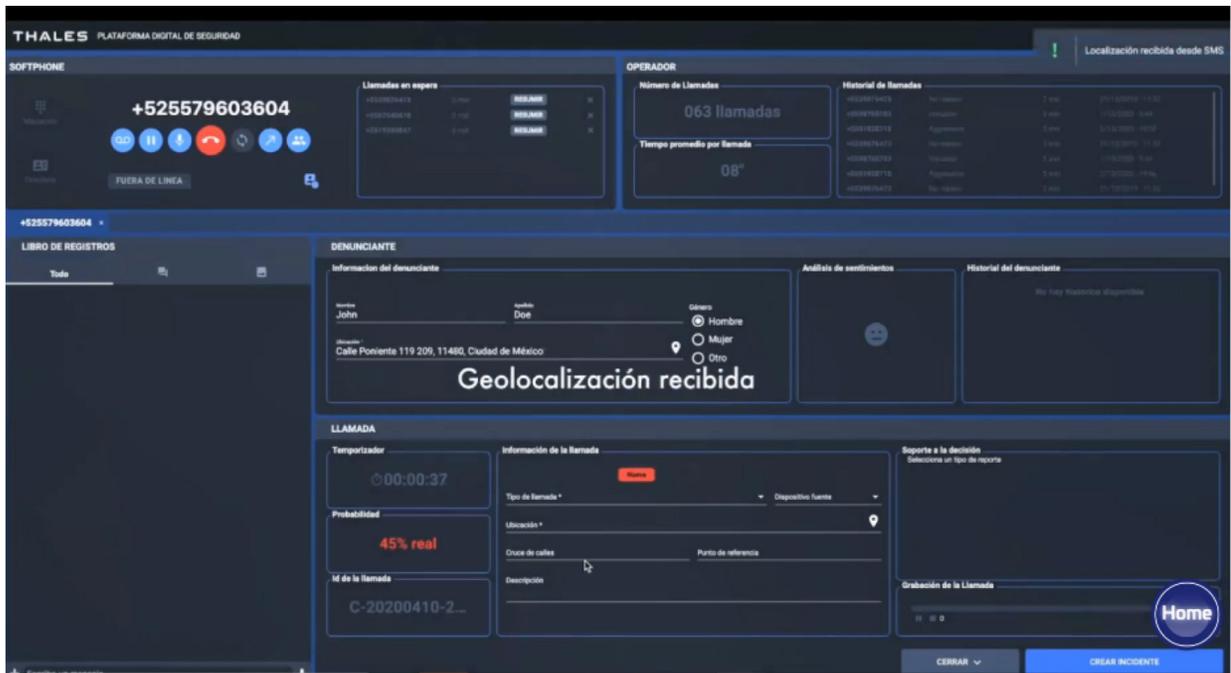


**LA CRISE DE LA COVID (VIDÉO PORTANT SUR LE DÉPLOIEMENT AU MEXIQUE)**

➤ *Exploitation de données de centre d'appel photo du panel :*



➤ Fournir des aides pour mieux qualifier les appels. On géolocalise



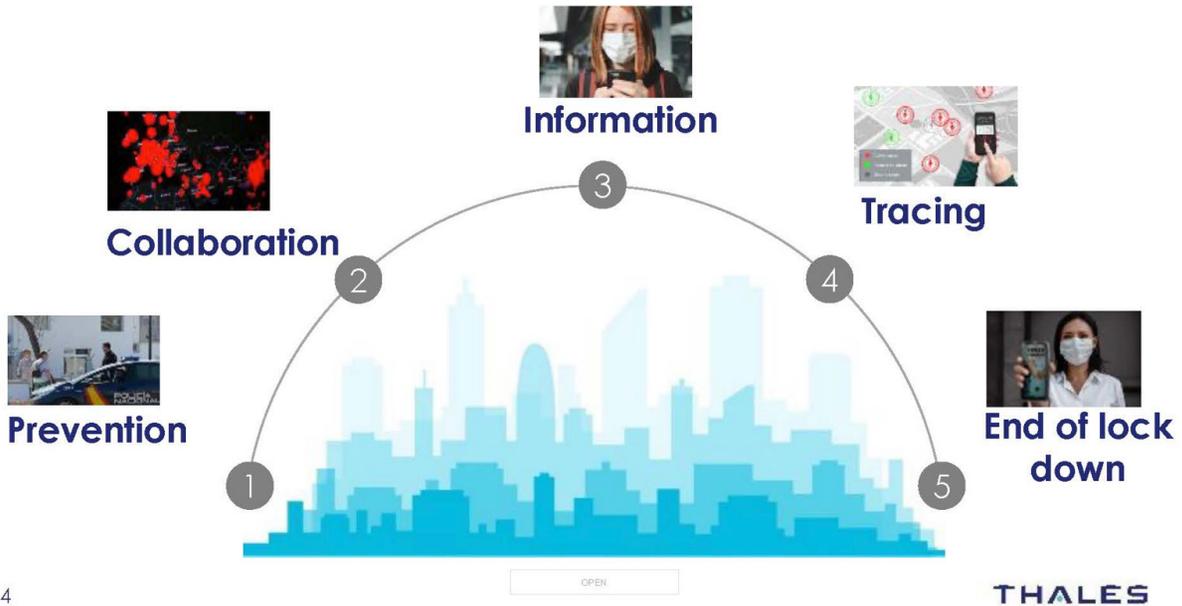
➤ On remonte ces informations dans la SDP, et on synthétise :



Cet outil a été développé pour la crise de la covid, sur base de la plate-forme SDP. On remonte des indicateurs par État.

## Challenges for Covid-19 crisis

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. ©Thales, 2015. All rights reserved.



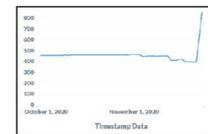
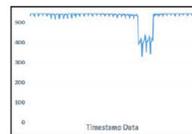
14

### ➤ Exemples de cas d'utilisation

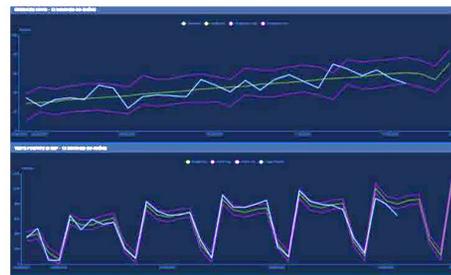
This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. ©Thales, 2015. All rights reserved.

#### USE CASE #1 - Automatic detection of data errors and anomalies

Alerts for Phone - Alert Quantified	Alerts for Phone Age - Alert Quantified	Alerts for Phone To - Alert Quantified	Alerts for Phone Cost - Alert Alert Quantified	Alerts for Phone Cost - Alert Alert Quantified	Alerts for Phone Cost - Alert Alert Quantified
Number of Alerts	Number of Alerts	Number of Alerts	Number of Alerts	Number of Alerts	Number of Alerts
0	0	0	0	0	0



#### USE CASE #2 - Weak signals and abnormal trends detection



15

Il faut savoir agréger de plus en plus de données pour la plate-forme et, d'abord, assurer une détection de problèmes et anomalies dans l'ingestion des données : c'est l'étape de validation des données.

*Doit-on modifier les données en cas d'anomalie reconnue ?*

**N**on, on choisit d'afficher les données sources, mais on qualifie les sources. On ne modifie pas les données on avertit celui qui fournit les données de ses problèmes. Ce choix est important car on peut toujours remonter des alertes non justifiées. Et donc corrompre des données en croyant les modifier à bon escient.

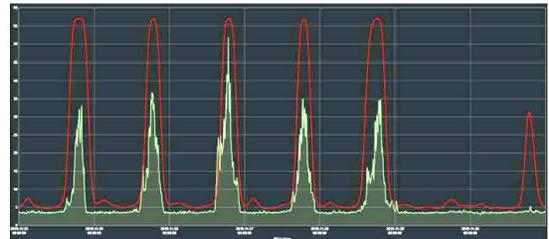
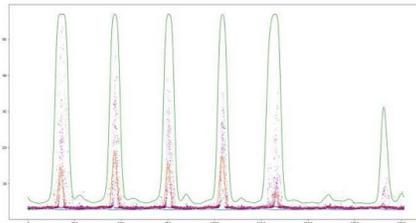
Dans le cas d'utilisation #2 on fait une détection de changement de dynamique (tendance) ou d'anomalie (sortie de fourchette de valeurs).

## USE CASES

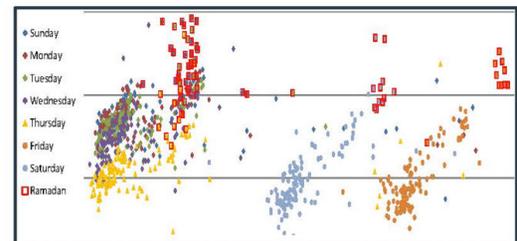
This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. - © Thales, 2015. All rights reserved.

### USE CASE #3 – Anomaly detection in large volume of data

(> 10M points per day)



### USE CASE #4 – Identify patterns and similarities in the data



16

VM FR – Oct 2019 v2

OPEN

THALES

Ici on fait des détections fines sur un gros volume de données, ce que permet le recours aux réseaux de neurones. Contrairement au cas #2, ici le volume de données permet de construire un modèle sans faire d'hypothèses sur la distribution statistique de la donnée, d'où une fiabilité accrue dans la détection d'anomalies.

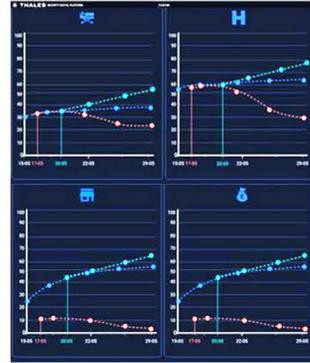
Le cas d'utilisation #4 est une visualisation des données permettant d'identifier des patrons (*patterns*), et les paramètres externes corrélés. Ceci est ici illustré sur des données de trafic en fonction du type de jour. Ici on utilise des modèles de prédiction selon différentes hypothèses pour comparer différents scénarios, en support à l'aide à la décision.

## USE CASES

**USE CASE #5** – Forecast based on prediction models or scenarios



**USE CASE #6** – Impact analysis by comparing different scenarios, such as reopening schools, restaurants or public transportation



This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales, 2019. All rights reserved.

17

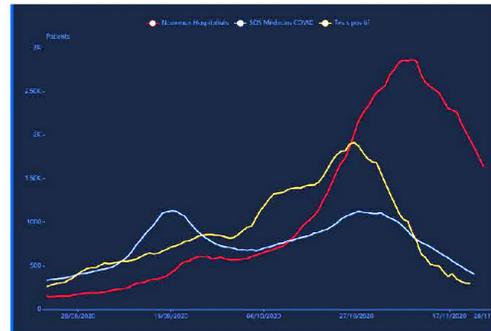
OPEN

THALES

## USE CASES

**USE CASE #7** – Identify time correlation between indicators.

For example, which indicator is the most reliable predictor of the future rate of new patients in intensive care?



This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales, 2019. All rights reserved.

18

OPEN

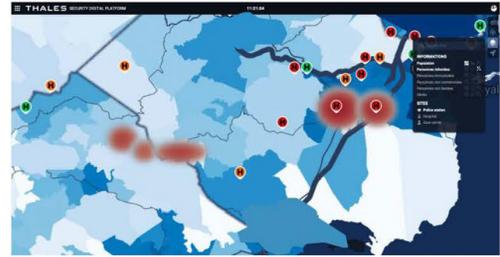
THALES

Dans le cas d'utilisation #7, on cherche à mesurer les corrélations ente indicateurs ou informations. On peut ainsi se servir des indicateurs rapides pour prédire les indicateurs lents, ou identifier les indicateurs les plus pertinents pour anticiper l'évolution de la situation sanitaire.

## USE CASES

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales 2015. All rights reserved.

**USE CASE #8** – Monitor contamination clusters



**USE CASE #9** – Monitor large population movements (week ends, holidays)



19

VM FR – Oct 2019 v2

OPEN

THALES

Dans les cas d'utilisation #8 et #9, on surveille les clusters et les mouvements de population. Toutes ces données peuvent être intégrées à un modèle compartimental. C'est la capacité d'intégration de la plate-forme qui est essentielle. On fait la même chose pour les frontières : comparaison entre pays pour comparer les différentes mesures.

### ➤ Ville intelligente et sécurité aéroport

## Commercial solutions for states/cities/sites

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales 2015. All rights reserved.

<div data-bbox="310 1340 565 1425" data-label="Section-Header"> <h3>SDP-Crisis</h3> </div> <div data-bbox="310 1448 565 1598" data-label="Figure"> </div> <div data-bbox="321 1616 553 1671" data-label="Text"> <p><b>Massive data ingestion &amp; analysis</b></p> </div> <div data-bbox="321 1682 553 1779" data-label="List-Group"> <ul style="list-style-type: none"> <li>➤ Detect abnormal body temperature</li> <li>➤ Multi-site monitoring</li> </ul> </div>	<div data-bbox="607 1340 857 1425" data-label="Section-Header"> <h3>Security Box Thermo</h3> </div> <div data-bbox="607 1448 857 1598" data-label="Figure"> </div> <div data-bbox="643 1616 1065 1646" data-label="Text"> <p><b>Plug &amp; Play Solution for incident detection</b></p> </div> <div data-bbox="581 1671 860 1765" data-label="List-Group"> <ul style="list-style-type: none"> <li>➤ Statistical Dashboards</li> <li>➤ Situation awareness with stock &amp; resources</li> </ul> </div>	<div data-bbox="862 1340 1112 1425" data-label="Section-Header"> <h3>Security Box Lockdown</h3> </div> <div data-bbox="862 1448 1112 1598" data-label="Image"> </div> <div data-bbox="878 1671 1117 1788" data-label="List-Group"> <ul style="list-style-type: none"> <li>➤ Photo + QR code encrypted PKI</li> <li>➤ Enable authorities to control circulation</li> </ul> </div>	<div data-bbox="1159 1340 1414 1425" data-label="Section-Header"> <h3>COVID-ID</h3> </div> <div data-bbox="1159 1448 1414 1598" data-label="Image"> </div> <div data-bbox="1192 1616 1430 1671" data-label="Text"> <p><b>Secure Management of individual virus status</b></p> </div> <div data-bbox="1182 1682 1446 1779" data-label="List-Group"> <ul style="list-style-type: none"> <li>➤ Detect (un)lockdown rules violation</li> <li>➤ Leverage legacy CCTV</li> </ul> </div>
---	--	--	---

Ready for COVID19 crisis and future needs

21

VM FR – Oct 2019 v2

OPEN

THALES

## Commercial solutions for airports

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales, 2015. All rights reserved.

### 1 Occupancy Control

- Thales InFlow Occupancy Control**
- Passenger Flow Management System
  - Implements boarding pass gate
  - Machine learning-based algorithm
  - Powered by AI



### 2 Ensure Social Distancing

- Thales InFlow Wearable**
- Connected wearable
  - Powered by UWB/BLE type tag



Social Distancing

### 3 Crowd Detection

- Thales SafeLand Pax Protector**
- Thermal cameras
  - Geo-localized alarms
  - Overcrowd areas detection



### 4 Life Tracking

- Thales InFlow LiTracking**
- Lidar/BLE sensors
  - Powered by AI

22

Vile F1 - Oct 2019 v2

OPEN

THALES

## B. ÉLÉMENTS TECHNIQUES ET FACTEURS CLÉS

### KEY SUCCESS FACTORS

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales. © Thales, 2015. All rights reserved.



Aware

Avoid blind spot in your decision making



Precise

Consult synthetic & intuitive crisis status & evolution



Predictive

Simulate and anticipate crisis and its evolution



Collaborative

Better communicate & share internally & externally?

9

Vile F1 - Oct 2019 v2

OPEN

THALES

## LES DONNÉES

Notre expérience est que souvent la difficulté est d'obtenir les données. Les propriétaires sont arc-boutés dessus. Il y a un mélange de difficultés, y compris politiques. De façon générale, les données sont un élément de pouvoir.

### ➤ Identifier les sources et collecter les données

Un point clé est la capacité à récupérer les diverses sources :

- entrée manuelle des données quand on n'a rien d'autre. SDP offre donc des interfaces facilitant cette saisie. Elles sont adaptées à la diversité des utilisateurs, voir § suivant ;
- nous collectons les données institutionnelles et étatiques, voir à ce sujet le § suivant ;
- les réseaux sociaux sont intéressants ; les appels d'urgence aussi.

### ➤ Structurer et organiser la donnée

- Il y a des degrés de maturité très différents. Certains services de ministères sont bien organisés et préparés pour la collecte et l'exploitation des données, d'autres pas. Certains clients nous demandent même si nous savons traiter des données non structurées (document pdf, planches Powerpoint...) ;
- Depuis peu, SDP offre en plus de l'analyse de sources ouvertes (réseaux sociaux), de la détection de signaux faibles, et des analyses de sentiment ;
- SDP offre une capacité à récupérer toutes ces données et les met à disposition des clients pour qu'ils construisent eux-mêmes leurs indicateurs à partir des données (aide précieuse pour la prise de décision).

### ➤ Valider les données

On a besoin d'effectuer une détection de problèmes et anomalies lors de l'ingestion des données : on parle de validation des données. C'est important, car on peut remonter des alertes non justifiées.

On ne rectifie pas les données jugées corrompues. On se contente de les repérer et de les annoter avec l'indication de leur source. On qualifie les sources. On avertit l'utilisateur, ainsi que celui qui fournit les données de ces problèmes.

Le cas des données de l'Insee est assez typique des données étatiques. Elles ont un certain retard (qui peut être variable, exemple des décès). Beaucoup de sources de données sont comme ça. C'est une question gênante. Il faut qu'en permanence on récupère tout l'historique des changements *a posteriori*. On sait qu'à court-terme on travaille avec des données biaisées, mais ce biais est mesurable et peut être indiqué.

On compare aussi différentes sources censées donner le même indicateur, pour estimer le niveau de confiance, qui est la chose importante qui nous est demandée. On qualifie la confiance

dans l'indicateur fourni par ce type de comparaison ainsi que la réputation des sources.

### **HÉTÉROGÉNÉITÉ ET SYNTHÈSE**

L'aspect collaboration est fondamental. On partage des vues partielles des données représentées selon les rôles et cultures des utilisateurs.

L'hétérogénéité culturelle est un frein. Ainsi, les différents ministères utilisent des acronymes différents, et le même acronyme peut avoir un sens différent selon les ministères. Ceci crée des difficultés. À chaque niveau de décision différent il y a une façon d'appréhender les choses différente, avec le cas extrême de l'État belge. En France des arrêtés municipaux relatifs à la pandémie ont été en contradiction avec des arrêtés préfectoraux, eux-mêmes en contradiction avec celles de l'État. On a dû tout repenser pour la covid : édicter des règles pour les coiffeurs, les écoles... qui constituent autant de contextes différents.

On a donc offert des interfaces intuitives et simplifiées pour permettre les échanges, avec, en particulier, des interfaces de type smartphone, pour éviter la courbe d'apprentissage.

Partager le même outil au niveau horizontal permet de mieux partager l'information et la compréhension. Si vous ne voyez pas ce que voit l'autre, vous avez du mal à vous mettre d'accord. On essaie de faire tomber les cloisons entre ces diverses administrations ou ministères.

### **MODÈLES ET LEUR INTÉGRATION**

Fournir des informations synthétiques pour l'aide à la prise de décision a été la demande principale pour la crise de la covid. SDP offre des modèles pour jouer des scénarios et en prédire les effets. On fait une partie de nos modèles, mais on en importe aussi beaucoup.

#### ➤ **Modèles compartimentaux, leur estimation et la détection de changements**

On cherche à comparer les résultats des modèles. On fait donc l'exercice de validation des modèles en permanence. On utilise pour cela des modèles de type série temporelle qui marchent bien en horizon court tant qu'il n'existe pas de grand changement (par exemple, de type confinement à déconfinement). Pour prendre en compte ces changements on a recours à des modèles compartimentaux (avec plusieurs modes, chacun portant des paramètres différents). La difficulté est de bien en caler les paramètres. On les complète par des algorithmes de détection de changement, qui nous permettent entre autres de détecter des changements de mode, impliquant des changements dans la dynamique épidémiologique.

#### ➤ **Machine learning**

##### **Images radiographiques**

On travaille sur la détection des images radiographiques (3,5 millions) pour identifier une origine covid, avec des réseaux convolutionnels. Ces images peuvent ensuite être collectées, affichées pour les utilisateurs et utilisées pour des algorithmes de *machine learning* répondant à d'autres questions, plus spécifiques. On a déployé cette infrastructure en trois mois y com-

pris l'apprentissage. Dans ce type de développement, tout va bien jusqu'au PoC<sup>91</sup>, passer à l'opérationnel est le vrai challenge.

*On n'a pas toujours 3 Millions d'images pour démarrer. Comment faites-vous ?*

On a une base algorithmique très variée, plus large que le *machine learning*. Il y a beaucoup de cas où on a de l'information et des connaissances métier et on fait mieux avec de la modélisation. On a donc souvent des algorithmes très compacts qui tournent sur des machines petites. Parce qu'on a des modèles (que l'on paramètre) et la connaissance qui va avec. Dans le cas d'une pandémie, les modèles compartimentaux sont des équivalents permettant de modéliser la propagation de la pandémie. Mais ces modèles nécessitent de nombreux paramètres non directement mesurables. On peut caler ces paramètres sur la réalité par des approches de type algorithmes génétiques, par exemple.

### AGILITÉ

La plate-forme SDP permet une configuration rapide d'un outil pour un client particulier. On travaille en mode agile pour livrer rapidement les versions successives à nos clients (tous les quinze jours). Il faut aussi avoir l'agilité permettant de se déployer sur divers environnements du client. Tout ceci est une compétence de fond dans le groupe. Ainsi, pour un client nous avons fourni un premier aperçu de la situation trois jours après avoir reçu les données. Généralement on met les choses en place en quelques jours. On déploie des solutions très rapidement c'est surtout du paramétrage, un peu comme le déploiement d'un VPN (réseau privé virtuel) en télécommunications.

### INFRASTRUCTURE POUR L'INTÉGRATION DE COMPOSANTS (DONNÉES, MODÈLES, SERVICES)

La plate-forme SDP est basée sur des standards qui nous permettent d'intégrer des données et des composants offrant divers services. Traitements en *streaming* ou *batch*. Analytics Studio est l'outil de développement des modèles. L'agilité est fondamentale (déploiement tous les quinze jours). Il faut pouvoir déployer sur des plates-formes de façon agnostique vis-à-vis de l'architecture cible. Pour y parvenir, l'architecture de SDP est innovante et évolutive : utilise des micro-services, des services de stockage de données pour ne pas être limité dans l'évolution d'échelle. SDP offre des présentations utilisant des cartographies. Argo est utilisé pour faire des suites de traitement en batch génériques, et JupyterHub pour les suites de traitements de modélisation, qui permettent de faire collaborer des outils et composants. On utilise un modèle commun et générique pour la structuration des données, avec des contraintes imposées concernant le format et la méthode de stockage. On distingue sources brutes, résultats de traitements, métadonnées...

91 Proof of Concept.

### **LA SALLE D'ANALYSE**

L'exemple de salle montrée dans une des premières planches comporte beaucoup d'écrans et d'opérateurs. Elle est typiquement faite pour la gestion des incidents, problème assez décentralisé pouvant être conduit de manière décentralisée par beaucoup d'opérateurs. D'autres crises (décision de type « confinement » dans la covid) sont des décisions beaucoup plus centralisées. Thales est souvent acteur dans la définition de ces rôles opérationnels.

Le Conseil de défense était à l'origine une cellule réduite. Pour opérer la gestion de la crise de la covid, c'est devenu un centre interministériel (travail de type militaire journalier pour faire les points et remonter les éléments synthétisés). La gestion de la crise covid est bipolaire : santé et conseil de défense, ce qui complique la chose.

SDP propose d'appuyer ces travaux humains par de la technologie pour aider et accélérer ce type de travail. Ceci existe déjà depuis longtemps dans le secteur de la défense, avec le même degré de complexité : coopération stratégique-tactique en France et pour l'Otan.

## ANNEXE 2 — ENTRETIEN AVEC DASSAULT SYSTÈMES

*(9 Octobre 2020)*

*Notes de Albert Benveniste et Laurent Gouzènes, revues par Patrick Johnson*

### ➤ *Participants*

#### **Académie des technologies**

**ALBERT BENVENISTE**

**LAURENT GOUZÈNES (invité)**

**PATRICK JOHNSON**

**GÉRARD ROUCAIROL**

#### **Dassault Systèmes**

**PATRICK JOHNSON**

**MAHEL ABAAB-FOURNIAL**

**VINCENT SIESS**

## PRÉSENTATION DE LA STRATÉGIE « JUMENTAUM NUMÉRIQUE POUR LA GESTION DE CRISE », PATRICK JOHNSON

La société Dassault Systèmes regroupe 200 000 personnes dont 8 000 ingénieurs. elle est effectivement engagé dans la gestion de crise au niveau national, avec plusieurs équipes (Affaires globales, R&D, solutions), des objectifs, et le développement d'une plate-forme outillée. Les outils correspondants sont d'ores et déjà en service aujourd'hui sur la gestion de la pandémie.

Nous allons commencer par montrer nos travaux visant à héberger des modèles d'épidémiologie sur notre plate-forme nous sommes à l'état de l'art en ce qui concerne ce type de modélisation, tant sur les modèles de la famille SIR (à base d'équations différentielles), que sur les modèles multi-agents. Comme dans tous les développements de modélisation de la covid, la calibration des modèles est une réelle difficulté, dès que les modèles cherchent à monter en finesse et en fidélité (modèles SIR compartimentaux, par tranche d'âge ou profil de personne, ou par mode de vie et c'est encore plus vrai quand on cherche à calibrer des modèles multi-agents).

Mais ce qui nous importe le plus dans notre développement de plate-forme, c'est de prendre en compte l'aspect multidimensionnel (économie, industrie, logistique, société, régalienn...). C'est ce qui manque aujourd'hui car le besoin est la compréhension de l'ensemble des dimensions impactées face à une crise. Cela a nécessité des modélisations intégratives permettant une optimisation multicritères. C'est ce qu'on a essayé de faire sur cette plate-forme : on vise en fait un jumeau numérique d'une collectivité (communauté, région, échelon national) pour un système de gestion de crise. De ce point de vue, on continue la vision de la société en passant du jumeau numérique du produit, à celui de la production, à celui de l'entreprise à celui de la ville et des territoires pour arriver aux régions et aux pays.

Pour faire cela, Dassault Systèmes dispose de trois atouts uniques :

1. La notion de **jumeau numérique du pilotage étatique**, le « cockpit ».
2. **Une plate-forme qui supporte l'innovation ouverte**. Car la gestion de crise n'incombe pas à l'étatique seulement on l'a donc ouverte aux hôpitaux, à des scientifiques et à toute partie « prenante ». La plate-forme est donc ouverte et propose une innovation ouverte (par exemple plusieurs équipes ont travaillé sur des systèmes innovants de respirateurs ou même de modélisation en *drug design* : communauté [OPEN COVID](#)).
3. **Une plate-forme pour les essais cliniques du monde** : 50 % des essais cliniques commerciaux mondiaux sont réalisés en utilisant la plate-forme [Medidata](#) de gestion des essais cliniques et pharmacologiques et des tests d'équipements médicaux. Pour cela, Dassault Systèmes a racheté la société Medidata (6 milliards d'euros pour ce rachat) qui est leader mondial sur ce créneau. Tous les essais cliniques dont on a entendu parler concernant la covid-19 sont opérés sur cette plate-forme qui figure maintenant au sein de l'offre Dassault Systèmes.

En utilisant la même plate-forme, Dassault Systèmes a ainsi fait des jumeaux numériques pour de nombreux secteurs : ingénierie des systèmes et produits, production, ville numérique, et maintenant décision étatique.

## PRÉSENTATION DE LA PLATE-FORME DASSAULT SYSTÈMES 3DEXPERIENCE — MAHEL ABAAB-FOURNIAL

### GÉNÉRALITÉS

Nous avons monté une équipe dédiée pour cette situation COVID pour voir comment la plateforme [3DEXPERIENCE](#) de Dassault Systèmes permettrait de mieux répondre à la crise, et aussi de mieux se préparer aux futures crises. On a développé pour cela une solution à destination de la puissance publique dans cinq secteurs :

- approche cockpit pour les gouvernements : aide à la prise de décision, suivi de l'exécution de cette décision, réconciliation des différents acteurs pour former une intelligence collective
- santé, avec le volet [Medidata](#) et son élargissement ;
- sécurisation des lieux publics : sûreté, circulation de l'air, lieux de spectacle, aéroports...
- économie : résilience, réallouer les ressources existantes, capacités à transformer un système logistique...
- support à l'éducation continue avec une préparation à l'adaptation aux ruptures, en particulier pour l'ingénierie.

Nous avons une brochure et des *position papers* sur tout cela.

### APPROCHE COCKPIT

Une démonstration live du cockpit est faite. La démonstration porte sur la région grand-est, qui portait 15 % de la population infectée au printemps. Notre plateforme a accompagné les autorités grand-est depuis mai 2020. Tout ce que vous voyez aujourd'hui dans cette démo est ou sera rapidement opérationnel. Vous voyez ici un cockpit santé, mais ce qui est couvert est plus large. Nous préparons la même chose pour la transition énergétique, pour la relance économique, etc.

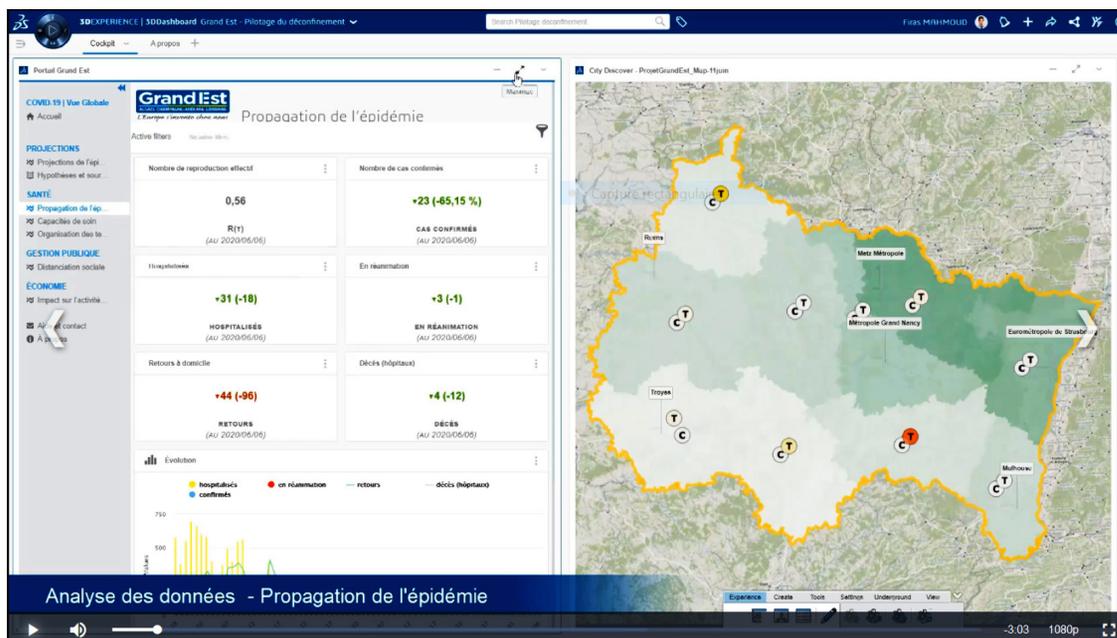
On a construit un observatoire de la donnée de santé en collaboration avec l'IHU Strasbourg et PRISM INESIA. La commande la région Grand Est auprès de PRISM était en substance « aidez-moi à protéger mieux ma population ». Le nom de code du service est [METEO-COVID](#). On y offre :

- une aide à l'interprétation de la donnée de santé. On fait une remise à jour journalière. On peut zoomer sur une région, sur une fenêtre temporelle ;
- des projections jusqu'à deux mois sur l'évolution de la crise ;
- du croisement de données pour interpréter des signaux faibles pour la relance économique. Cette dimension est importante car elle est plus profonde que l'aspect purement technologique Data Science. Elle permet à différentes structures (organismes, ministères) d'avoir une vue 360° ensemble.

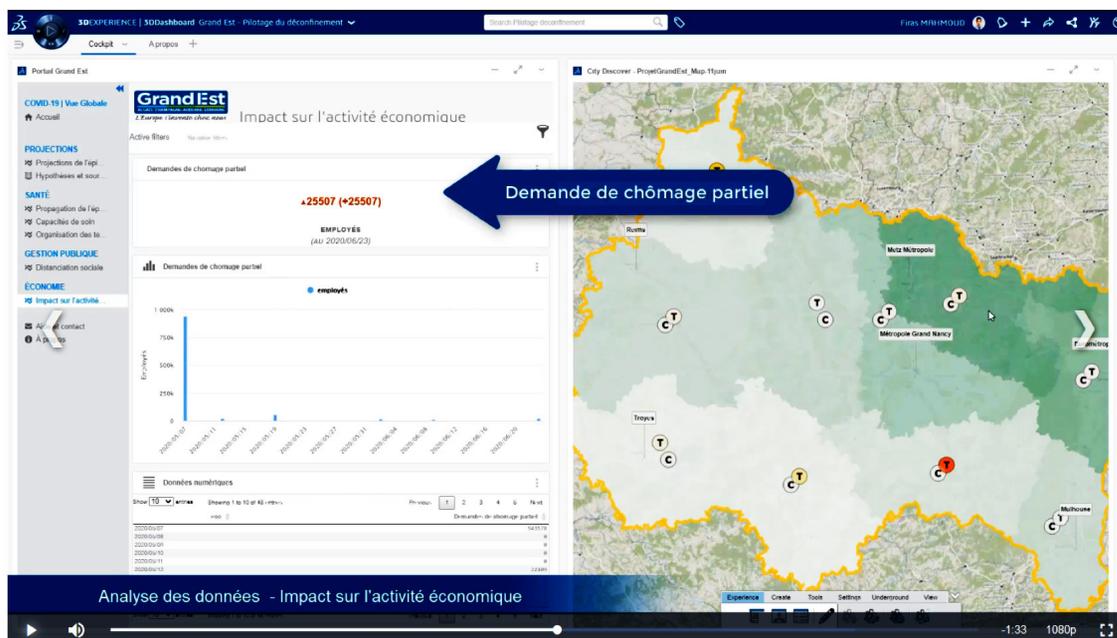
On fournit pour cela un tableau avec les divers indicateurs considérés. La région pourra y visualiser et identifier les clusters. Nous allons y intégrer les données provenant de SOS médecins et de l'ARS.

Sont impliqués dans le développement de cet observatoire Inesia : l'institut Pasteur, Inria, l'IHU Strasbourg, [Pulsi](#) (opérateur en communication grand-est). Au total, il y a plein d'acteurs impliqués qui forment un écosystème. Aujourd'hui la plate-forme a 300 utilisateurs. On montera jusqu'à 1500 utilisateurs.

➤ *Vue du cockpit de la région Grand-Est : carte active et indicateurs*



➤ *Vue du cockpit de la région Grand-Est : impact sur l'activité économique*



## QUESTIONS ET DISCUSSION

*LG : Vous avez évoqué des modèles finement compartimentés ? En avez-vous effectivement ? Par tranches ?* Aujourd'hui on inclut les tranches d'âge, les dynamiques de comportement. On veut surtout fournir un outil de description de modèle et les utilisateurs spécialistes vont pouvoir ajuster ces modèles, les recalibrer, etc. On a vu que les divers groupes de chercheurs proposent des modèles différents. Nous ne prenons pas parti, mais nous fournissons des outils qui permettent de prendre en compte les modèles au fur et à mesure de leurs évolutions.

*AB : quid de la gestion des modèles compartimentaux ? Comment peut-on rentrer tous les paramètres ?* On a une matrice qui décrit l'intensité des relations entre classes d'âge. Les territoires denses vont avoir une dynamique différente. Et quand on veut aller dans des choses plus fines, il faut aller dans des modèles agents pour décrire des populations d'individus. Par contre ces modèles agents sont plus coûteux que des modèles SIR, qui n'ont pas du tout démérité dans le suivi du COVID. Il faut donc faire des aller-retours entre ces deux modélisations.

*LG : peut-on diffuser de la peur pour faire se bien comporter la population ? Pouvez-vous mesurer indirectement le niveau de peur et vous en servir ?* Oui, on pourrait le faire sur des modèles agents. À la fin, toute modélisation pourrait proposer de prendre en compte certains comportements et leurs interprétations (par exemple la peur de sortir). Mais, attention, il faut se poser bien-sûr la question de la représentativité et de la calibration possible ou pas.

*LG : Avez-vous mis des compétences diversifiées dans vos équipes pour cela ?* Pas encore. Mais nous sommes de plus en plus en train d'ajouter des composantes *digital humanities*, qui pourraient en particulier traiter ça. Mais c'est le futur. Sans aller jusqu'à modéliser les différences culturelles, si on observe à l'issue des calibrations que les paramètres varient d'un pays à l'autre, on peut expliquer les différences.

*Quid des applications de traçage : il y a des acceptabilités très variables selon les cultures.* Ce serait intéressant de pronostiquer l'acceptation selon les cultures. Sur la modélisation par agents, on est en discussion avec l'Institut Luxembourgeois qui héberge des données de *mobile tracing*. On mobilise une équipe là-dessus. Sur cette dynamique internationale on verra sans doute les différences.

*Et sur l'organisation des tests PCR ?* Aujourd'hui on collecte les données et on surveille les capacités de test. Sur le volet socio-économique, on collecte les données relatives à la demande d'activité partielle, au chômage. Je peux zoomer et voir comment ça va à tel ou tel endroit. On intègre des croisements avec le taux de pauvreté ou d'autres paramètres. Évidemment, on peut faire de la data science là-dessus.

Mais au-delà, ce qui est important est de forcer la prise de décision collégiale et concertée entre ministères. La concertation est par définition difficile. Notre plate-forme est destinée à faciliter cela. C'est un incitateur de prise de décision collégiale, plus qu'un support de la data science (ces technologies y figurent, mais c'est la coordination qui est l'appui principal). On est sur un paradigme de coopération politique, au-delà de la seule épidémiologie. On cherche à créer une intelligence collective par circulation de l'information et traçabilité de l'échange d'information. On sait gérer finement qui voit quoi, et qui a accès à quoi (tout ceci remplace avantageusement les fameuses « notes » des collaborateurs).

L'enjeu de tout cela est de permettre de collaborer pour assister la prise de décision. On a une suite d'outils qui permet de croiser les données. On met à la disposition des modules simples qui permettent de croiser des données et de les corrélérer pour aider à la décision. C'est puissant pour des décideurs politiques (Imaginons par exemple Olivier Véran et Bruno Lemaire, croisant leurs données et en tirant leurs décisions en temps réel). C'est avec cet objectif en tête qu'on propose le jumeau numérique pour la décision politique : on veut permettre à tout un chacun de pouvoir comprendre le champ de l'autre, de croiser, corrélérer et synthétiser une vue 360 multi-dimensionnelle en temps réel mais simple à comprendre. La *user experience* doit aussi être la plus intuitive pour cela.

*GR : c'est un invariant de ces plates-formes de permettre de dé-silotiser. Je vais être audité par la commission Bothorel sur les données.*

## ÉLÉMENTS TECHNIQUES D'INGÉNIERIE SYSTÈME, VINCENT SIESS

Deux démonstrations sont faites : l'une sur modélisation/simulation épidémiologique, l'autre sur modélisation/simulation de système de soin national.

On voit une interface [Dymola](#) au sein de CATIA Systèmes. Cette approche système est très générale, on l'a utilisée en biologie, pour la modélisation du cancer, pour les dispositifs médicaux, pour le génie en bioprocédé, l'économie, etc...

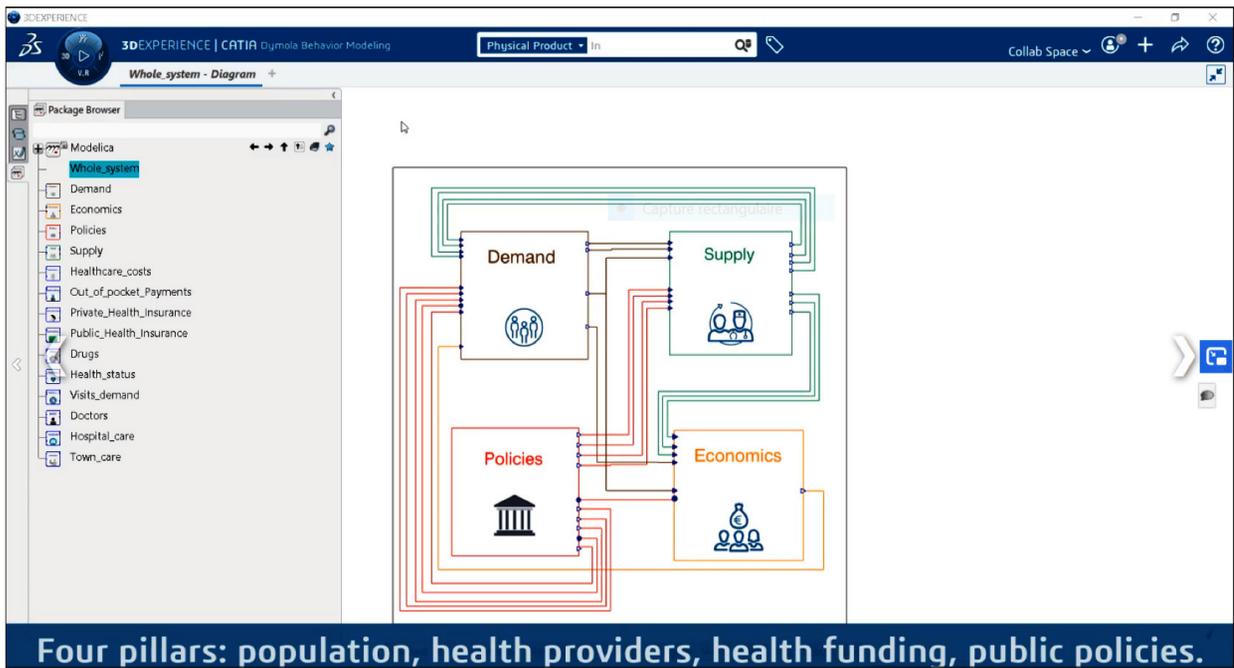
Note d'AB sur Dymola :

*« C'est l'outil Dymola qui sert d'infrastructure de modélisation, pour tous les aspects où une modélisation mérite d'être développée : épidémiologie, mais aussi le cockpit pour les décideurs politiques. L'outil Dymola est un outil qui permet de représenter les évolutions au cours du temps de systèmes régis par des dynamiques continues (équations différentielles...) ou discrètes (agents). L'outil offre un langage noyau textuel qui permet d'éditer finement des modèles mathématiques, la possibilité de définir des palettes de modèles prédéfinis et de les réutiliser, et une interface graphique qui rend l'outil rapidement utilisable par un non-technicien du domaine. Un code de simulation est ensuite généré, qui permet de jouer les scénarios et d'obtenir les projections. Ce type de modélisation peut être associé à des techniques d'assimilation de données ou d'apprentissage, pour calibrer les modèles sur des données lorsque c'est nécessaire. »*

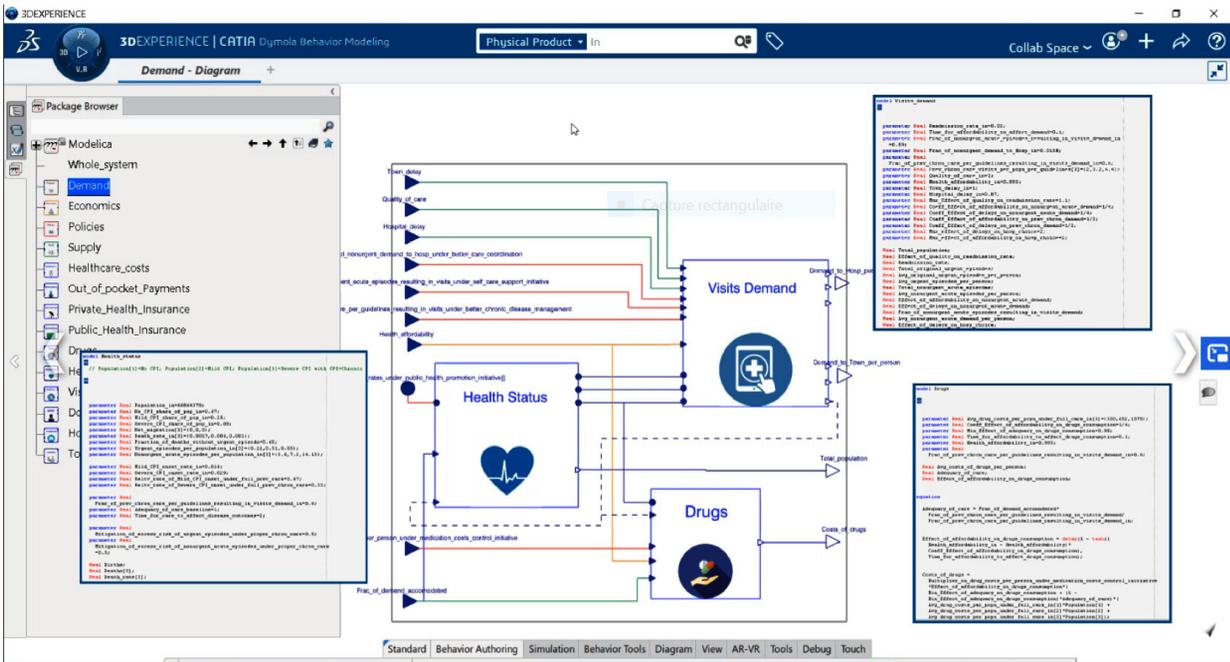
Pendant les premières semaines on n'avait pas de données. On ne savait pas qui était infecté, on connaissait seulement les décès. On a dû par conséquent se limiter à des modèles présentant une certaine topologie et faire des calibrations adaptées et moins robustes que si des données avaient été disponibles en plus grande quantité. Puis, il a fallu faire intervenir des compartiments qui codent les autres aspects (charge des hôpitaux...). On a donc rajouté ces variables. On a des compartiments SIR (modèles d'épidémiologie à base d'équations différentielles, ODE).

À côté des équations qui décrivent l'impact de l'épidémie sur la prise en charge des malades du coronavirus, nous avons également développé un modèle à l'échelle d'un système de santé complet (à plus long terme, pour de la prévision stratégique). Il existe très peu de modèles de cette envergure dans le monde un autre aux États-Unis ([ReThink Health](#)). Notre modèle a été calibré par rapport à des données publiques récupérées sur le site de l'État. Les données en bleu sont les données gov.fr et en rouge la calibration du modèle on constate une bonne fidélité. On estime et on affiche la sensibilité du modèle en fonction de la date de prise de décision. Ainsi, on sait aujourd'hui que, si l'on était entré en confinement une semaine plus tôt, on serait comme l'Allemagne en termes de bilan. Avec une semaine plus tard, on aurait aujourd'hui 50 000 morts. Les modèles illustrent très bien ça. Selon les pays ou les sociétés, l'acceptabilité plus ou moins grande et la confiance dans les pouvoirs publics ont, en partie, fait que les décisions ont été prises plus ou moins tôt ou tard. On fournit aussi une étude de l'impact économique de telle ou telle décision. Cela permet de gérer les compromis entre économie et santé.

➤ *National Health System Model: vue globale*

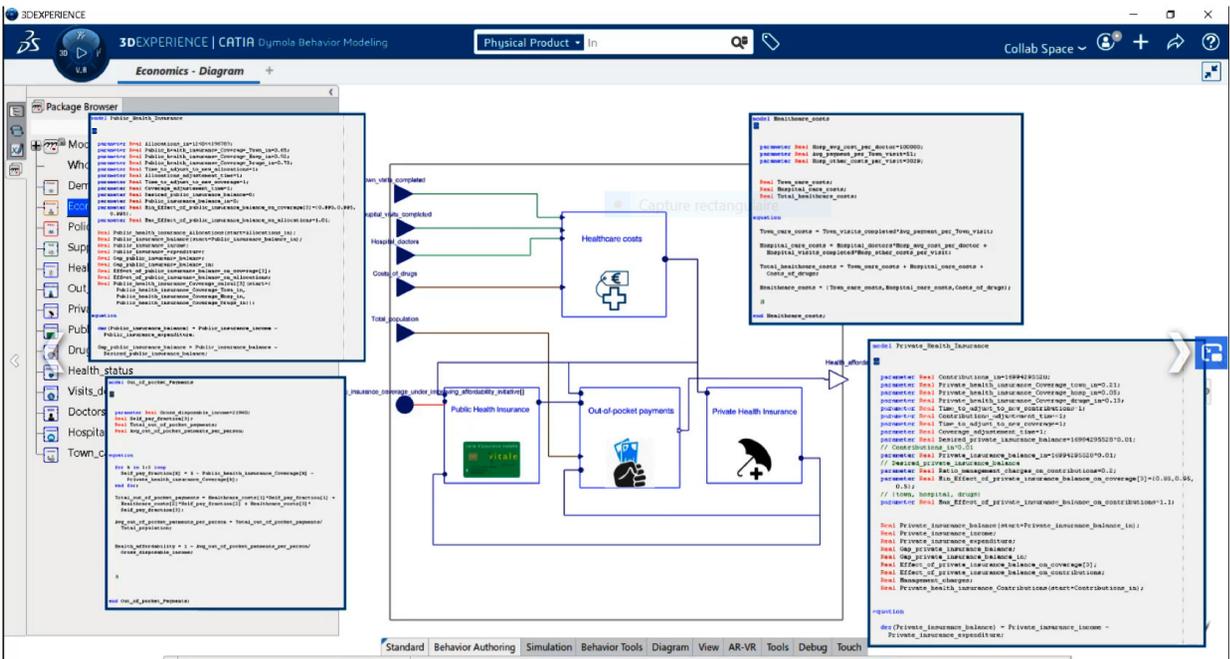


➤ National Health System Model: zoom sur la boîte « demande »



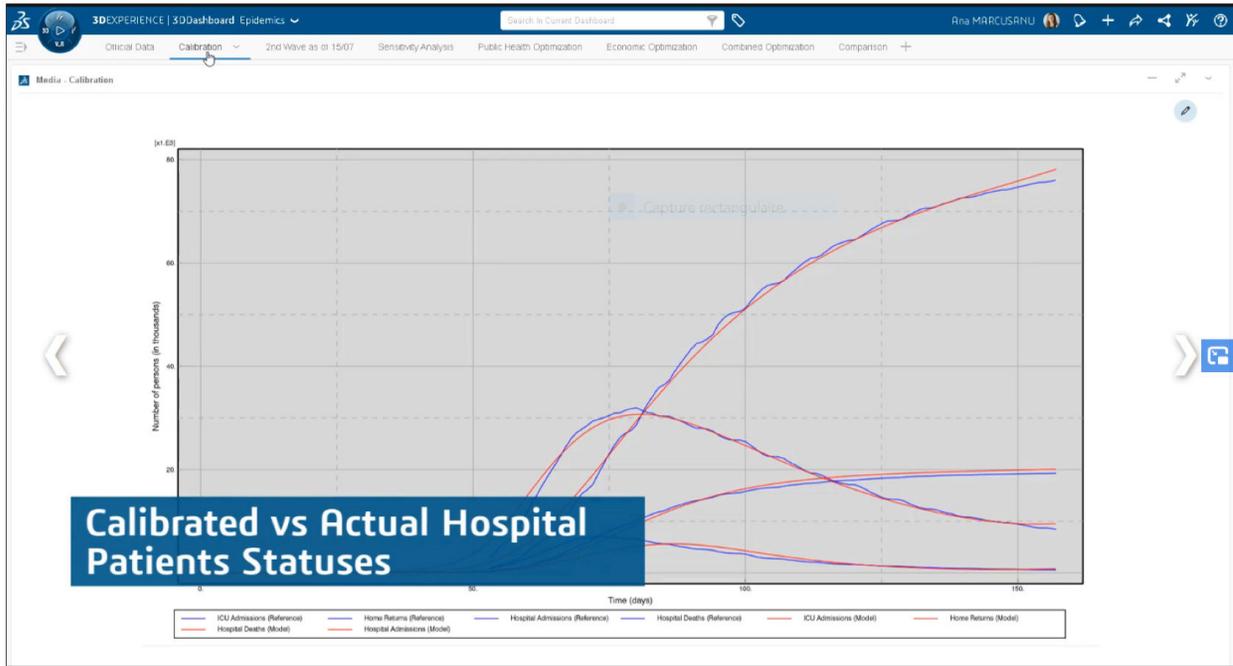
On voit apparaître le premier niveau de ce modèle « demande ». On voit aussi, sur les bords, du texte du langage Modelica qui est celui de l'outil Dymola. Les boîtes ne sont donc pas seulement des dessins, mais des modèles effectivement simulables.

➤ National Health System Model: zoom sur la boîte « economics »



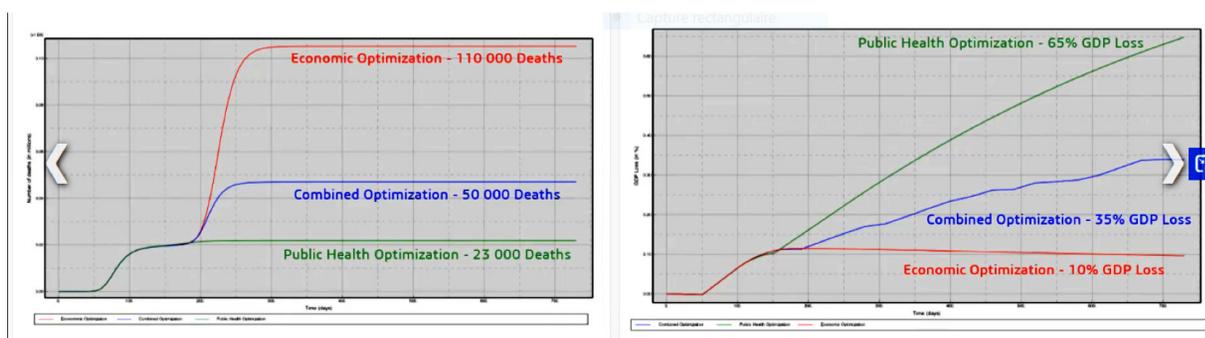
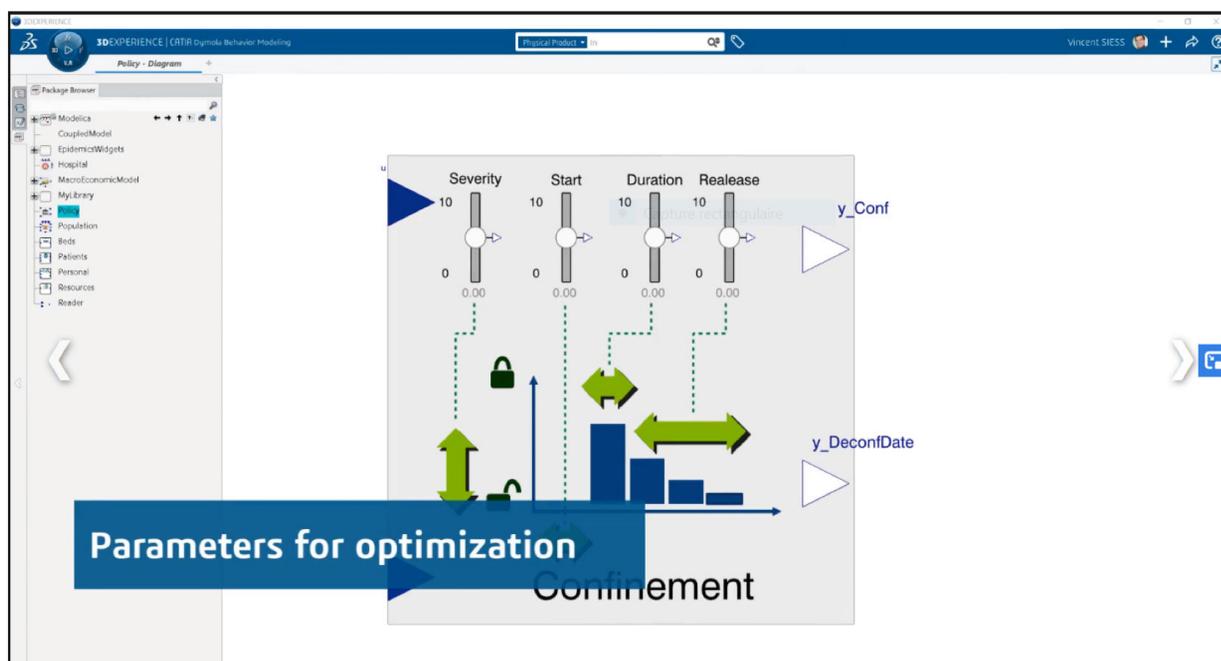
Plus généralement, Dassault Systèmes veut couvrir le monde du soin industriel (essais pharmacologiques) et le monde politique. On se pose la question de s'attaquer aux prévisions macroéconomiques dans le même esprit. On a commencé à calibrer ce modèle sur les données disponibles sur douze ans en macroéconomie. On note que la calibration reste difficile pour certaines variables, c'est en cours.

➤ **National Health System Model: zoom sur la fonction « calibration », résultats**



En rouge, les prévisions du modèle calibré en bleu les évolutions vraies constatées.

➤ **National Health System Model: zoom sur la fonction optimisation, en deux planches**



Une difficulté est la modélisation des politiques publiques c'est le point difficile en fait. On peut piloter les systèmes de santé, voir les effets des décisions de pilotage.

AB : *Quid du support de communication vis-à-vis du public ?* Il faut rendre les choses simples et compréhensibles. Faire que tout cela soit utilisable par le politique est une difficulté. On a traité ça. Sur la dissémination et la pédagogie grand-public, c'est une vraie question sur laquelle on devra réfléchir. Cela fait écho au *portail citoyen* qu'on développe pour les *smart cities* (sortie prévue en fin d'année). Comment peut-on utiliser ça pour communiquer ? On a bien ça en tête. Ça induit des mécaniques différentes. On y réfléchit. Je crois beaucoup au comparatif : par exemple, pour les protocoles d'usage des applications de traçage, si on lui dit comment ça marche en Angleterre, Mme Michu comprend, mais si on lui explique comment marche *StopCovid* elle ne comprend rien.

## ANNEXE 3 — ENTRETIEN AVEC PIERRE HAREN, CAUSALITYLINK

*Notes de Albert Benveniste et Laurent Gouzènes, revues par Pierre Haren.*

### ➤ **Participants**

#### **Académie des technologies**

ALBERT BENVENISTE

LAURENT GOUZÈNES

GÉRARD ROUCAIROL (excusé)

#### **Causality Link**

PIERRE HAREN

ERIC JENSEN

CHRISTIAN DEUTSCH

#### **Thales**

MAKREM DRIDI

#### **Dassault Systèmes**

PATRICK JOHNSON

### ➤ **Codes couleur:**

- Points importants soulignés par Causality Link
- Points importants soulignés par le GT

### ➤ **Autres codes:**

Comme pour les annexes précédentes :

- Les questions sont indiquées en italique et les réponses en roman
- Les auteurs des questions sont indiqués par leurs initiales

## SYNTHÈSE

### OBJET DE L'ANALYSE : COMPRENDRE LES LIENS DE CAUSALITÉ

Pour la plupart des questions qui se posent à notre société, c'est un enjeu majeur que de comprendre quels sont les liens de causalité entre événements, faits, tendances, etc. Ces liens de causalité sont un élément majeur pour comprendre les défis auxquelles nous devons faire face, ainsi que les actions qu'il est judicieux d'entreprendre.

### APPROCHE DE CAUSALITY LINK ET OUTIL

L'IA peut contribuer à ça. Le premier point important à noter est que les approches statistiques (dont *machine learning* et *deep learning*) ne permettent pas d'inférer des causalités ; le statisticien Judea Pearl explique très bien pourquoi. Si l'on veut ajouter des éléments de causalité dans un modèle statistique, il faut le faire explicitement (on peut ensuite valider ou invalider cette hypothèse de causalité statistiquement).

Causality Link a choisi d'inférer ces liens de causalité par agrégation d'opinions humaines, collectées par des agrégateurs de contenu payant (dans des flux considérés comme jouissant d'un minimum d'autorité, pas Twitter...). La collecte des liens de causalité se fait

- en continu, par analyses des flux de documents ;
- en utilisant des techniques d'analyse de la langue naturelle (après traduction de 23 langues vers l'anglais), assistées d'ontologies.

Pour cela on repère des syntagmes spécifiques et on les analyse en tenant compte du contexte associé (il existe des sous-entendus dans ces éléments de phrase). Les ontologies sont adaptées continuellement, en parallèle avec leur utilisation, par apprentissage. Causality Link utilise des techniques dites de *text mining*, non issues de modélisations statistiques. La gestion de ces gros volumes de documents et de données est effectuée dans l'outil public [Elasticsearch](#).<sup>92</sup> Chaque lien de causalité est ensuite étiqueté par des informations agrégées à partir des attributs quantitatifs collectés dans les syntagmes associés à ce lien (pour un lien de type « due to », on agrège les paires {cause, conséquence} associées) on tire ainsi des indicateurs agrégés, que l'on peut afficher sur un tableau de bord doté d'outils flexibles de navigation.

### CIBLES : DE L'ANALYSE FINANCIÈRE À L'IMPACT DE LA COVID-19 SUR L'ÉCONOMIE

La cible initiale, pour quoi l'outil est conçu et dimensionné, est la population des analystes financiers. L'outil leur apporte des moyens de prévision des marchés en fonction d'informations collectées, relatives à tel ou tel secteur économique ou acteur économique. La présentation a utilisé comme illustration les duopoles Apple/Huawei et Tesla/NIO.

Causality Link a appliqué la même démarche aux flux relatifs à la covid-19, avec pour objet l'analyse des impacts sur l'économie. On a ainsi un exemple de complément aux simples modélisations de type épidémiologique.

92 Des techniques d'apprentissage statistiques pour des analyses sémantiques de bases de documents sont utilisées par d'autres auteurs, mais le credo de Causality Link est qu'ils font mieux avec leurs méthodes symboliques.

## OBSERVATIONS ET QUESTIONS

Rechercher les liens de causalité exprimés par des humains en tant qu'opinion, les identifier et les agréger est une démarche de portée très générale. C'est clairement un outil clé dans la gestion de crise, pour établir les liens de cause à effet, les mesurer, pour pouvoir ensuite les contrôler, en particulier de la part des pouvoirs publics.

L'outil développé par Causality Link est un exemple remarquable de l'information que l'on peut extraire et synthétiser lorsque l'on dispose d'une base de documents d'échelle mondiale, sur un ou un ensemble de domaines. Il existe maintenant une grande variété d'outils automatisant ce type de démarche, les plus connus étant sans doute les outils d'exploration de bases de brevets ou de bases de données bibliographiques. Son argument fort est l'extraction de causalités par identification et agrégation d'opinions de ce type, exprimées dans des documents collectés sur divers flux — la plupart des autres outils se concentrent sur l'étiquetage des bases à l'aide de descripteurs, qui servent ensuite à répondre à des requêtes, comme pour les bases de brevets.

Si l'objet de l'analyse (liens de causalité) ne peut être inféré dans le cadre d'une modélisation statistique, l'analyse de type langue naturelle qui doit être développée pour faire l'identification et l'agrégation de ces causalités peut être envisagée de diverses manières : *text mining* symbolique (comme dans Causality Link) ou *deep learning* via le [word embedding](#). D'un point de vue général dans le domaine de la langue naturelle en IA, la compétition entre ces deux approches est encore ouverte aujourd'hui.

Attention, il n'existe aucune « vérité terrain » à quoi puissent être confrontées les informations remontées par cette démarche aux utilisateurs. La pertinence des réponses ne peut être évaluée que par jugement humain (une autre opinion, après tout c'est moral). À ce titre, ce type d'outil ne peut aucunement être considéré comme un « algorithme prenant une décision à notre place », ça demeure un assistant. En conséquence, la qualité de l'interface de navigation dans la multiplicité des informations (documents bruts, ontologies, et indicateurs agrégés) est un point essentiel.

## CAUSALITY LINK

Il y a quatre ans on ne parlait que de *machine learning* (2012-2016). Il se disait : avec *Watson on déploie et le système s'améliore*. Mais ça ne marche pas comme ça. Le futur n'est pas le *deep learning* tout seul, mais un mélange {statistique + symbolique}. En pratique, les gens de *machine learning* ne sont pas bons dans le domaine dont ils traitent les données, ils sont bons dans leur technique. La question qui nous est posée : comment agréger le savoir de foules immenses ? Le but premier de Causality Link est d'appliquer une telle démarche à la compréhension des marchés financiers.

## Causality Link: three converging forces

### 1. Mix harmoniously Symbolic AI and Statistic AI

- Symbolic AI: readable code that models conceptual thinking
- Statistic AI: automatic machine learning from lots of data

### 2. Apply the “wisdom of crowds” concepts to data AND knowledge (= causality)

### 3. Build automatically a model of the forces acting on the financial markets

- Worldwide
- All public companies
- All commodities



Confidential – Copyright Causality Link 2020

3



Confidential – Copyright Causality Link 2020

3

## The Company

### 1. Causality Link created in Utah in May 2016

### 2. Raised \$3M from friends and family

Ken Nickerson, technical founder of PDT

Tom Strat, launched the DARPA Grand Challenge in 2004

### 3. Donated 5% of the shares to the Toulouse School of Economics foundation

Jean Tirole, Nobel Prize of Economics on the Advisory Board

Three joint research projects underway

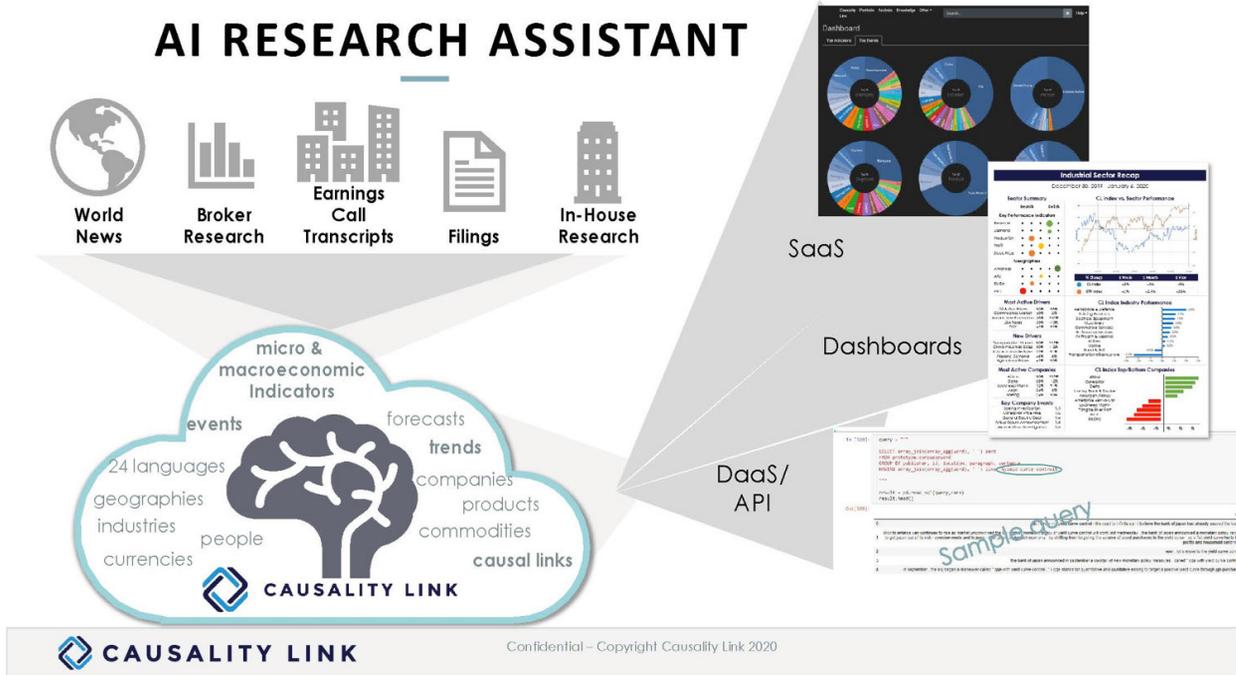
### 4. Raised \$5M in Series A in March 2020

### 5. Signed first large NY-based hedge fund customer in March 2020

“Causality Link’s focus on causal links or “the why” is clearly the next huge step forward in the use of AI systems for the investment research process.”

Michael Mayhew Integrity Research 10/15/2020

Jean Tirole est membre de l’Advisory Board il vient quelques jours par an. On fait la recherche du « why ? » (liens de causalité). Concernant le personnel, nous sommes 3 en France, 1 en Uruguay, les autres sont aux États-Unis. On est 12 en tout. [On fait de l’IA qui agrège la connaissance humaine sur les liens de causalité.](#)



**CAUSALITY LINK**

Confidential – Copyright Causality Link 2020

**Core Software technologies :**

- AWS based distributed architecture ;
- Prolog-based NLP (Natural Language Processing) on top of Spacy ;
- Elastic Search for content and index management ;
- Athena (AWS) data lake & query management ;
- Jupyter notebooks and BI tools for data science developments.

On capte des flux de textes choisis pour être intelligents (pas Twitter...). On a acheté les accès aux flux Naviga et à d'autres, qui publient des collections de textes et documents. On a 93 millions de documents dans notre système. On en tire des *indicateurs*, des événements, des *tendances*, et des *liens de causalité*. On a un SaaS (System as a Service) temps-réel qui répond à des requêtes.

On maintient une grosse ontologie dont l'ajustement est de plus en plus automatique par un mécanisme d'auto-apprentissage en opération. Le NLP établit la cohérence entre textes et ontologies. (On a dans nos effectifs deux experts mondiaux en NLP, l'un d'entre eux ayant déposé plus de 300 brevets en NLP pour IBM.)

Une difficulté centrale est d'arriver à gérer une trentaine de modules et composants sur étagère qui évoluent très vite, et d'assurer la cohérence de l'ensemble. En tant qu'outil, nous avons réussi une excellente mise en œuvre très performante de AWS alors que la gestion des ressources associées est complexe.

Les liens sont établis par des méthodes statistiques, ce qui leur confère de la robustesse.

## DÉMONSTRATION, ILLUSTRATION SUR LA COMPÉTITION HUAWEI/APPLE

Les 93 millions de textes arrivent par flux continu à haut rythme. On est dans un temps-réel assez exigeant. On a des millions de liens de causalité, des millions de concepts. On a plein de langues. On lit le chinois, l'arabe. On traduit tout en anglais.

On reconnaît les liens de causalité par des groupes de mots comme, par exemple, « due to ». On utilise des ontologies pour faciliter l'interprétation et éviter les interprétations qui sont des non-sens (par exemple, selon le contexte, une « amélioration » peut être une « augmentation » ou une « diminution »). Par hyperliens, on peut remonter directement des concepts aux textes qui les ont générés. On a donc une traçabilité complète. Un indicateur est constitué d'un KPI complété par des contextes. L'agrégation des indicateurs associés est un point performant.

**Exemple** : suivi des portefeuilles d'actions. Le cas de NIO (compétiteur de Tesla en Chine). On analyse les tendances sur le revenu de NIO qui se trouvent expliquées à partir des causalités identifiées dans les textes analysés.

*PJ* : de quel domaine relèvent les ontologies pour ce cas ? Il s'agit de l'interprétation que les gens font de l'activité des entreprises. C'est donc une ontologie spéciale pour ça.

*PJ* : Aujourd'hui je cherche au CNRS les types de collaboration, les types de brevets, etc. Faudrait-il forger une nouvelle ontologie pour ça ? On a développé ça récemment sur l'ESG (Environmental Social and Governance) on a sorti une bonne ontologie là-dessus en deux mois. Nous faisons ESG + finance, on découvre donc des liens entre les deux. On dépend des documents que les flux nous fournissent.

Je vous montre l'ontologie ESG c'est une arborescence chaque boîte de l'ontologie a ses paramètres. C'est construit semi-automatiquement, à l'aide d'un apprentissage en ligne en boucle fermée par une sorte de *bootstrap*. Les nouveautés à introduire dans l'ontologie sont proposées automatiquement, mais corrigées ensuite à la main. On absorbe aussi les contenus Wikipedia pour mettre à jour nos ontologies.

On montre le modèle causal de NIO. Plus on zoome sur ces graphes de causalité, plus c'est entaché d'erreurs.

*MD* : peut-on donner un indice de confiance aux différentes sources ? On est capables de noter les *publishers* par les qualités de leurs prévisions. Mais quand on a commencé à faire ça, ça a créé de la panique chez nos clients car on les notait. On a donc décidé de devenir plus discrets et on a mis la pédale douce. Les utilisateurs veulent souvent mélanger les flux extérieurs analysés par nous avec leurs flux intérieurs, à quoi ils accordent une valeur spéciale. Notre outil permet d'exploiter les signaux avant-coureurs des tendances.

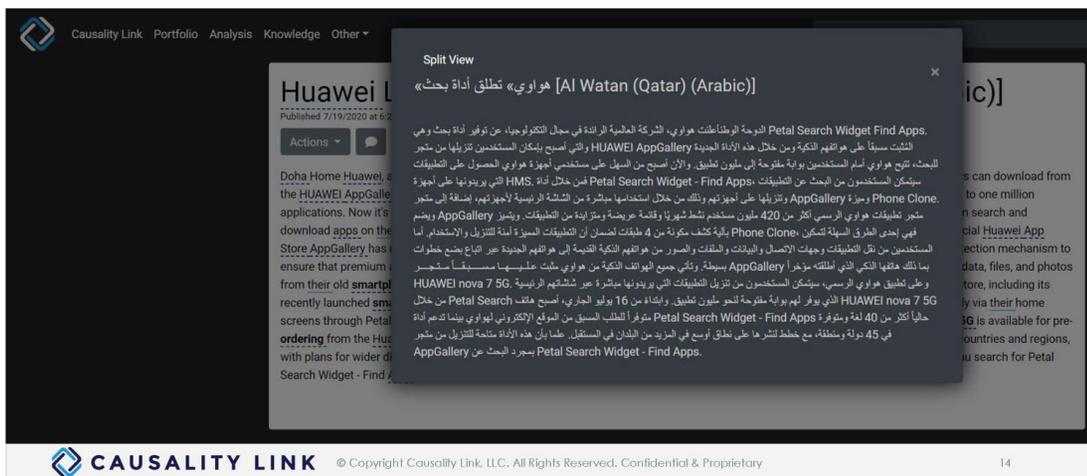
➤ Le « technological decoupling » :

## Apple / Huawei viewed by Causality Link

- 1- We are able to read news in Arabic and Chinese talking about Huawei
- 2- The “technological decoupling” announced by Xi on 10/29 was easily predictable in July 2020
- 3- Huawei / China is trying hard to replace the US technology in the Arab world
- 4- Huawei will increasingly put pressure on Apple in China
- 5- Same thing will happen to Tesla vs NIO, LI and a few other EV Chinese manufacturers

Nous lisons en ce moment les nouvelles en arabe. À propos de la Chine, Xi a annoncé un *technological decoupling*. Ça va se traduire en une bataille d'enfer entre Apple et Huawei. Idem pour Tesla vs. NIO dans le véhicule électrique. Huawei cherche à remplacer toutes les apps US sur les téléphones mobiles. Tous ces textes montrent l'évolution des choses entre Chine et pays arabes, avec les efforts de pénétration des Chinois. Voici un texte en Arabe :

### We read news in Arabic and Chinese talking about Huawei



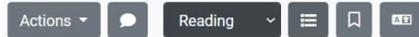
The screenshot shows a news article in Arabic from 'Al Watan (Qatar) (Arabic)'. The article title is « تطلق أداة بحث » هواوي. The text discusses Huawei's AppGallery and Petal Search Widget. The interface includes a 'Split View' window and a navigation bar at the top with 'Causality Link', 'Portfolio', 'Analysis', 'Knowledge', and 'Other'.

et l'analyse faite sur la traduction en anglais :

## We read news in Arabic and Chinese talking about Huawei

### Huawei Launches Search Tool [Al Watan (Qatar) (Arabic)]

Published 7/19/2020 at 6:25:39 PM



Doha Home Huawei, a world leader in technology, announced the provision of a search tool, Petal Search Widget Find Apps, which users can download from the HUAWEI AppGallery store already installed on their smartphones. Through this new search tool, Huawei offers users an open portal to one million applications. Now it's easy for Huawei users to get the apps they want on HMS devices. With Petal Search Widget - Find Apps, users can search and download apps on their devices using them directly from their home screen, the AppGallery Store and the Phone Clone feature. The official Huawei App Store AppGallery has more than 420 million active users per month and a wide and growing list of applications. It features a 4-layer detection mechanism to ensure that premium applications are safe for download and use. Phone Clone is an easy way to enable users to move apps, contacts, data, files, and photos from their old smartphones to their new phones with just a few simple steps. All Huawei smartphones come on top of the AppGallery store, including its recently launched smartphone HUAWEI nova 7 5G. On the official Huawei app, users will be able to download the apps they want directly via their home screens through Petal Search, which provides them with an open portal for about 1 million applications. As of July 16, HUAWEI nova 7 5G is available for pre-ordering from the Huawei website while Petal Search Widget - Find Apps now supports more than 40 languages and is available in 45 countries and regions, with plans for wider dissemination in more countries in the future. This tool is available for download from the AppGallery store once you search for Petal Search Widget - Find Apps.

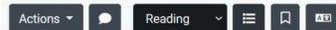
LG : quels outils de traduction utilisez-vous ? Google ou Amazon sont les meilleurs. En général c'est bien. Hindi à Anglais est le plus difficile. Notre dictionnaire de synonymes corrige des erreurs de ces traducteurs pour des termes importants (dont le coronavirus pour qui la traduction chinois → anglais est malheureuse). On a la chance que l'anglais soit une langue simple.

CD : comment as-tu sorti de Causality Link les textes que tu nous as montré ? C'est simplement en remontant, par hyperliens, depuis les liens de causalité exhibés vers les meilleurs textes qui expliquent les dits liens.

## Huawei / China trying to replace US technology in the Arab world

### Huawei Spark Up to 88% Eid Discounts

Published 7/20/2020 at 7:24:58 AM by المصدر: البيان الإلكتروني



Source:

— E-Statement Date:20 July 2020

After the great access of the first live digital event of discounts and as part of its comment to interact with local consumers by providing new offers, Huawei has unveiled the dates of HUAWEI AppGallery Eid Live Sale, the next edition of the UAE Live Consumer Event, which will come in time for the Eid holiday.

The live broadcast will take place over two days, Tuesday 21 July from 8:30pm on the official Huawei Arabic Facebook and YouTube pages, and Wednesday 22 July from 8:30pm on the official Huawei ic Facebook pages, YouTube and Mo Vlogs on Youtube.

Direct discounts are back active with a large number of the brand's latest products and its new flagship phone HUAWEI nova 7 5G.

Viewers will be able to get the finest deals for a limited period of time with up to 88% discount, huge and fast discounts and free prizes at home. They will also be able to interact with the live broadcast host on 21 July, footballer Adel Jamal and new hosts on 22 July: celebrity video bloggers Mo Vlogs and Lana Rose.

UAE residents at HUAWEI AppGallery Eid Live Sale will receive:

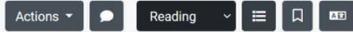
- Huge and fast discounts including purchase of 3 devices at AED 999: HUAWEI MatePad Pro 5G, HUAWEI Sound X, HUAWEI P40 Pro+ 5G and more 1
- Exclusive launch offer for 5G flagship HUAWEI nova 7 5G for AED 1,799 with free gift HUAWEI Watch GT 1
- Must not miss offers include HUAWEI Mate30 Pro 5G direct discounts at AED 2,399 (original price AED 3,899) and more 1
- Buy more and save more with extra discounts up to AED 600 when you buy more 1

— Spend AED 1,000 and get AED 50 off — Spend AED 2,000 and get AED 200 discount — Spend AED 5,000 and get AED 600 off - Won free prizes by sharing

## Huawei / China trying to replace US technology in the Arab world

### The Research and Knowledge Outreach Center and the China-Arab Studies Center for Reform and Development hold a joint seminar [Al Riyadh (Saudi Arabia)]

Published 7/16/2020 at 11:34:49 AM



The Research and Knowledge Outreach Center, in collaboration with the China-Arab Studies Center for Reform and Development of the Shanghai University of International Studies, organized yesterday a joint virtual seminar on the occasion of the 30th anniversary of the establishment of diplomatic relations between Saudi Arabia and China under the title "Review of the Past and Looking forward", with the participation of the Embassy of the Custodian of the Two Holy Mosques in Beijing and the Embassy of the People's Republic of China in Riyadh.

The director of the symposium, CEO of the China-Arab Studies Center for Reform and Development, Prof. Wang Guangda, spoke about the occasion and its establishment, which is of interest to both countries and calls for stopping and recalling many important and bright historical links in their relations.

After that, the opening session of the seminar began, attended by Ambassador of the Custodian of the Two Holy Mosques to the People's Republic of China, Turki bin Mohammed Al-Madhi, Ambassador of the China-Arab Cooperation Forum at the Ministry of Foreign Affairs of China, and former Ambassador to the Kingdom Li Chengwen.

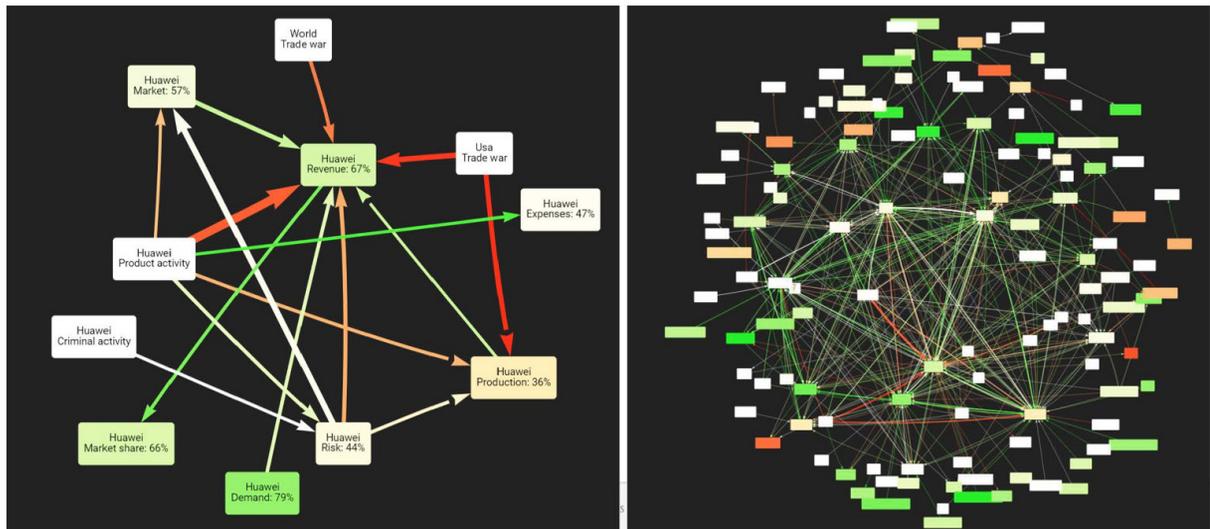
At the beginning of his speech, the last ambassador expressed his pleasure to be Ambassador of the Kingdom in Beijing during the 30th anniversary of the establishment of diplomatic relations between the two countries, which are celebrated and interested in as they represent friendship and respect between the two governments and peoples.

He recalled the efforts of the leading Saudi diplomatic personality, His Royal Highness Prince Saud Al-Faisal, who signed in 1990 the document establishing full diplomatic relations between Saudi Arabia and China, the exchange of ambassadors and the organization of meetings at the political, economic and youth levels, and others.

He stressed that the achievements of the two countries together are not measured in the past 30 years, but rather preceded and shortened it. Through their governments and people, relations have risen to business across all sectors.

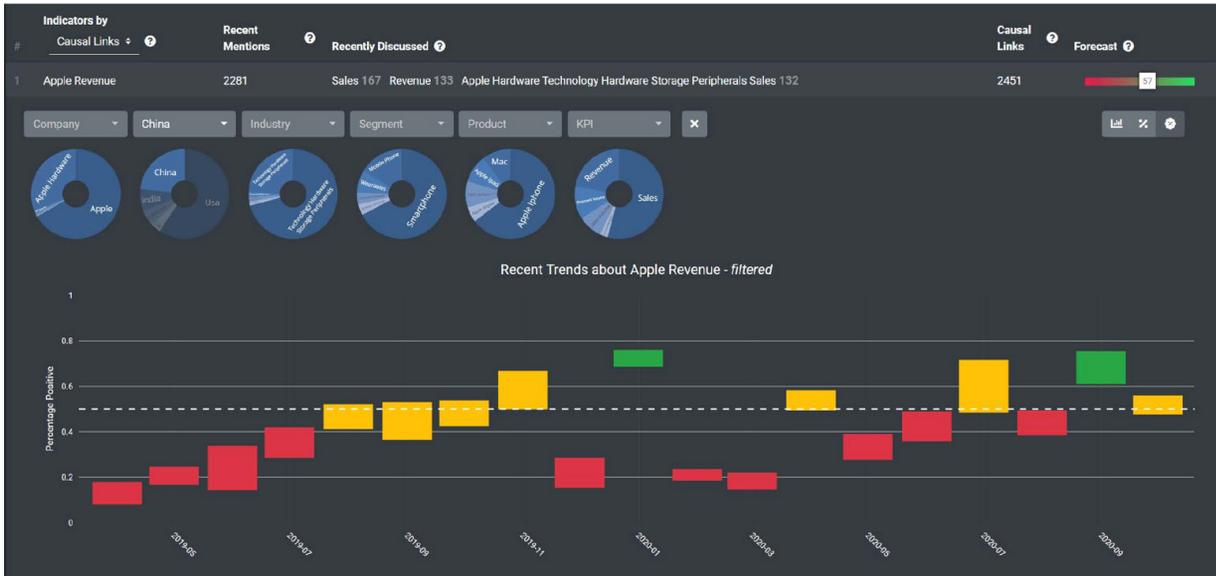
18

## Huawei will increasingly put pressure on Apple in China



Ces graphes montrent les liens de causalité trouvés pour Huawei, à deux résolutions différentes. On y voit des événements ou des tendances et les liens de causalité positifs (verts) ou négatifs (rouges) qui les relient.

## Huawei will increasingly put pressure on Apple in China



Notre outil sort donc des KPI (indicateurs) qui dessinent la situation de Apple telle qu'agrégée à partir de nos analyses (le rouge indique un indicateur comme ici le revenu dont le futur est prévu en décroissance a la date considérée).

Au total, la Chine est un énorme marché pour Apple et Tesla. La guerre commerciale États-Unis/Chine a causé le lancement du découplage, qui va favoriser les concurrents chinois de Apple et Tesla sur des marchés essentiels, en Chine et au-delà.

## COVID

Web site <https://coronavirus.causalitylink.com/>



On a vu le collapse arriver dès janvier 2020 pour l'aviation (*United*) ça a toutes sortes d'impacts prévisibles sur la consommation de pétrole, l'hôtellerie et la restauration.

### The market reacted 15 days later

#### Market Summary > Nasdaq Composite

INDEXNASDAQ: .IXIC

+ Follow

10,828.20 -357.39 (3.20%) ↓

Oct 30, 3:39 PM EDT · Disclaimer

1 day 5 days 1 month 6 months YTD 1 year 5 years Max

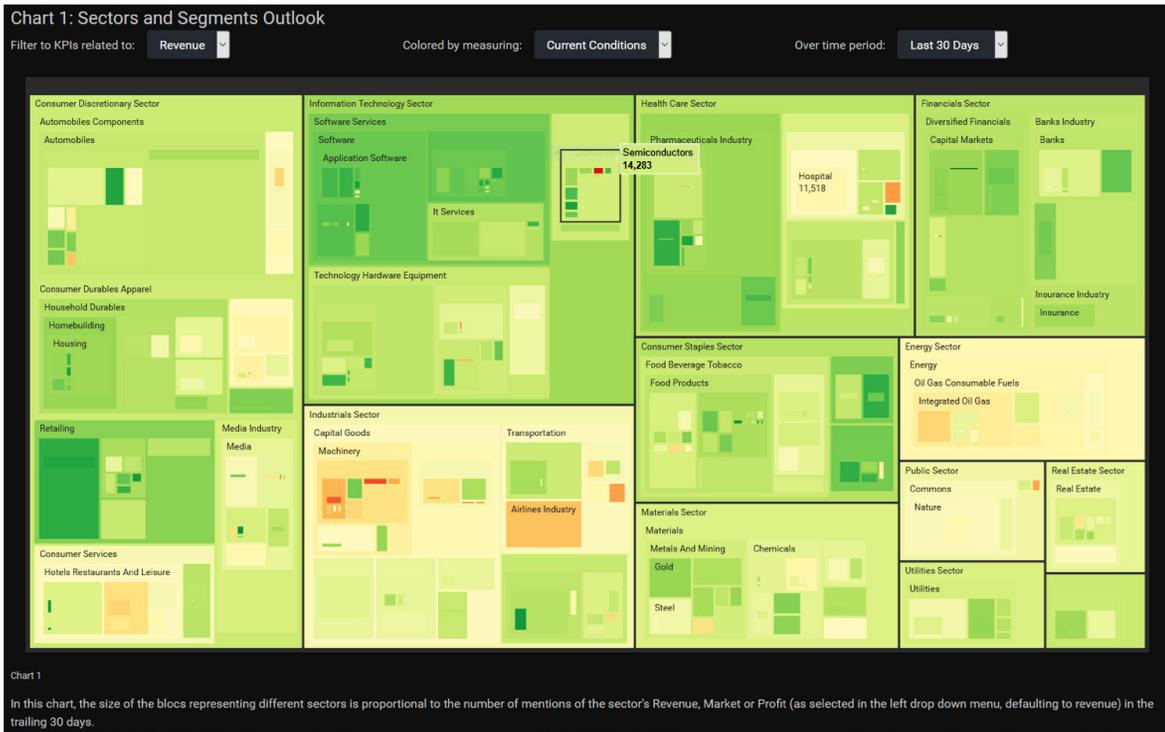


**CAUSALITY LINK**

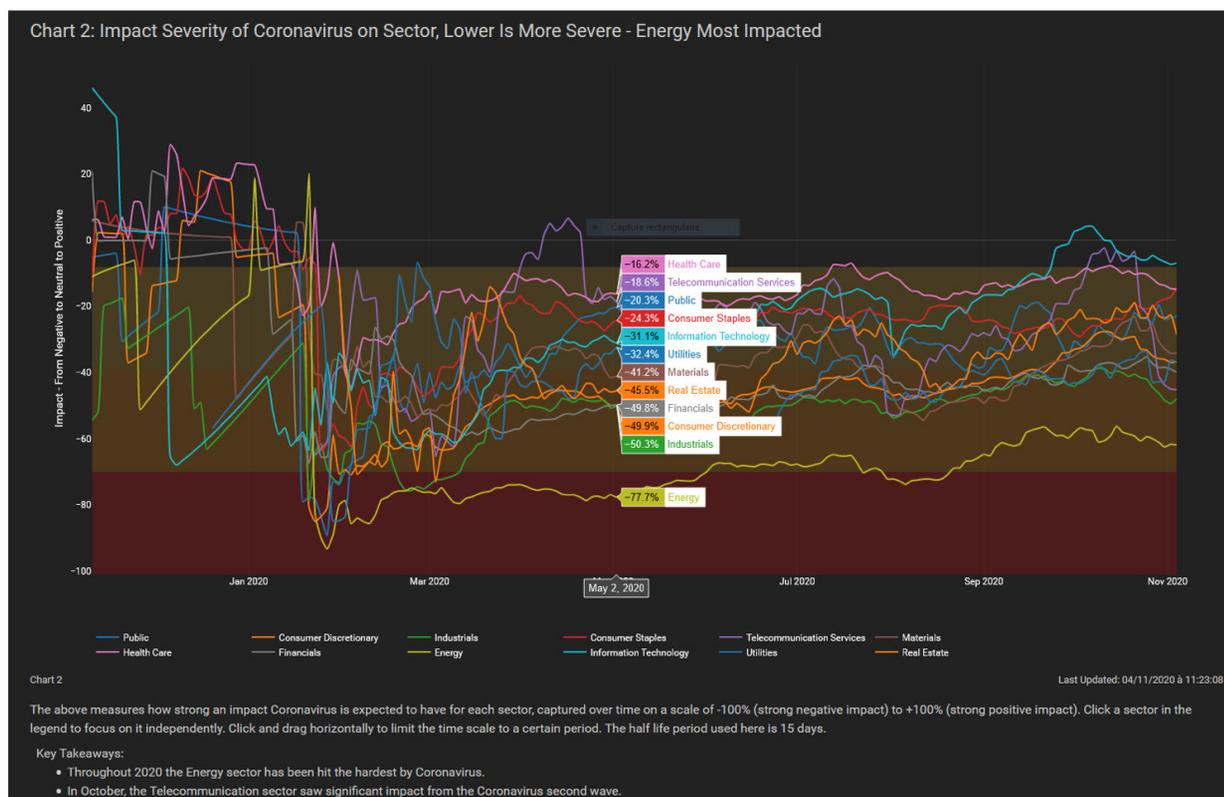
Confidential - Copyright Causality Link 2020

10

Le marché s'est écroulé le 19 février, c'est-à-dire presque trois semaines plus tard ! On a fait un tableau de bord. On y mesure toutes les prévisions d'évolutions pour tous les secteurs et les sous-secteurs. On a analysé l'évolution de tous les liens de causalité entre COVID et les diverses activités industrielles.



On voit ici les blocs représentant les différents secteurs industriels organisés hiérarchiquement. Chaque industrie a une couleur et l'axe vertical indique l'impact causal de la covid-19 sur la croissance future du secteur tel que perçu au cours du temps (presque partout négatif). On note l'impact de la seconde vague dans le monde.



On voit ici les impacts estimés pour les divers secteurs. Ceci est présenté à titre d'exemple d'usage possible de notre technologie. Nous sommes capables de faire des tableaux de bord sur la requête qu'on veut, sur n'importe laquelle des 40 000 entreprises ou des 3 000 secteurs que nous suivons.

LG : les flux ont-ils varié en quantité sur les divers sujets ? On a cette information, et la réponse est oui.

Noter qu'on a un mal incroyable à se souvenir de l'image du futur qu'on a eu dans le passé. C'est donc intéressant de pouvoir exhiber une image passée de ça. À la place on a coutume de réinterpréter l'idée du futur qu'on avait dans le passé. Mais nous pouvons revenir sur l'état exact de nos croyances en tout point dans le passé, puisqu'on a produit et archivé ces projections en temps-réel.

PJ : peux-tu nous en dire plus sur les statements causaux ? C'est délicat, comme le montre l'analyse des phrases comportant « despite ». Ces phrases cachent deux informations : 1/ ce qu'on attend normalement 2/ ce qui s'est en fait passé. Certaines phrases ne quantifient pas les évolutions, il nous faut donc boucher les trous avec des valeurs par défaut (qui font donc partie de l'ontologie). C'est donc compliqué et ça ne se fait pas par réseaux de neurones. Il

existe plein de formes de causalité : *due to* {*tendance*, conséquence *tendance*}. Attention, « *improved* » se traduit en + ou en – selon la sémantique de la quantité.

## COVID analysis

1. We tracked thousands of KPI for over 40,000 companies
2. We also tracked the causal links between COVID and these KPI
3. Observations:
  - 1- Government support has dulled the financial impact on companies
  - 2- The second wave will test the balance: Debt/Employment/GDP
  - 3- Comparing the triangle China/Europe/US is fascinating

C'est incroyable comment la dette a compensé l'impact de la covid dans la première phase. Mais on peut être plus inquiet sur la seconde vague.

L'exemple de Taiwan est saisissant (200 jours sans aucune contamination interne et un total sur 2020 de 7 morts dus à la covid). Mais ils ont tout mesuré et tout centralisé. Donc le GDP de Taiwan n'a pas baissé et ils n'ont pas de confinement. L'impact des aspects culturels est donc majeur.

On a maintenant un projet sur la comparaison globale du monde avant et après la covid. Et le différentiel de performance Chine/États-Unis est incalculable.

### CONCLUSIONS, LIENS, ET DISCUSSION GÉNÉRALE

#### ➤ *Conclusions from Causality Link*

*The future of AI lies in the ability to explain “why” and symbolic AI will stage a comeback. We will see the emergence of “collective brains” mediating human knowledge. The current COVID crisis must be understood in the context of:*

- *Trade-offs between globalization and decoupling ;*
- *Trade-offs between individual freedom and future capabilities ;*
- *Trade-offs between privacy and global performance.*

*Our societies face tough choices between individual and collective (future) well-being.*

Les gens sont en face d'une marée de documents, mais ne parviennent pas en extraire un savoir collectif.

➤ ***Le futur de l'IA est dans la causalité et l'IA symbolique va faire un retour.***

On va voir arriver des outils de médiation de savoir, qui vont agréger les choses et expliquer les relations. Les humains voient un maillon des causalités, mais notre outil agrège tous ces maillons. Exemple : on a pu mettre en évidence l'impact de la route de la soie Chinoise sur l'industrie automobile américaine par assèchement du marché de l'acier. L'analyse de ces causalités donne un avantage décisif en termes de timing sur les prévisions.

Je pense que la covid-19 va amener un incroyable impact sur le découplage technologique Chine-US qui amènera l'Europe à faire des choix difficiles. Dans le cas où l'Europe déciderait de refabriquer une grande partie de la technologie américaine (Google, Microsoft, Apple, Facebook, semi-conducteurs, télécommunications, matériel médical, etc...), il y aura une baisse de performance de l'économie EU vis-à-vis des États-Unis à cause du coût et du délai de remplacement. Idem pour la protection des données (cf. Taïwan). Toutes ces décisions auront un impact qui est rarement explicité lors de la décision.

➤ ***Liens communiqués par Pierre Haren***

<https://causalitylink.com/2020/02/28/a-causality-link-people-demand-supply-analysis-of-the-covid-19-outbreak/> : texte sur notre site Web que nous avons écrit fin février, et qui montre ce que nous pouvons déjà dire à cette date, avec les outils dont nous disposons alors.

<https://www.principles.com/the-changing-world-order/#chapter7> : texte de Ray Dalio, qui pointe sur un sujet plus vaste, mais qui prend encore plus de sens dans un monde qui traite la pandémie différemment selon les cultures.

➤ ***Discussion***

*AB : quid de la capacité à mettre à disposition des versions agrégées ou dégradées pour des usages plus larges et moins spécialistes ?* Aujourd'hui on est petits, on se focalise donc sur l'aide aux traders en prévision et notre outil est carrossé pour ça. Quand on aura plus de sous, on fera plus, et sans doute ça.

Mais le truc le plus fascinant est l'ESG (*Environmental Social and Governance*). La croyance générale actuelle est que plus tu fais d'ESG moins tu gagnes de l'argent. Cependant, nous détectons que certaines boîtes commencent à faire de l'argent avec. La covid a créé une prise de conscience de l'humanité sur les facteurs environnementaux. D'où l'importance de l'ESG. À court terme on peut aider à la prise de conscience en faisant de l'ESG bien fait.

*LG : saurais-tu contribuer aux grands débats du genre « nucléaire ou pas » ?* Il s'agit d'un débat dont les éléments principaux sont techniques. Or, nous ne sommes pas dans la vérité technique, mais dans la perception. Ce qui serait intéressant serait de coupler les jeux selon Yves Caseau pour la compétition (ex : GATES utilisé pour le pricing des licences de bandes radio chez Bouygues Tel) engendrés à partir de nos graphes de causalité.

*AB : quid de validation statistique a posteriori à la Judea Pearl ? On a essayé, on a la techno sous-jacente, mais on n'a pas terminé ce genre de travail. Donc, oui, on peut faire des validations a posteriori, mais c'est dur et ça paie peu, on a donc peu fait ça.*



## ANNEXE 4 — ENTRETIEN AVEC ALEX FLEISCHER ET ALAIN CHABRIER IBM

(3 Novembre 2020)

Notes de Albert Benveniste et Laurent Gouzènes, revues par Alex Fleischer.

### ➤ *Participants*

#### Académie des technologies

ALBERT BENVENISTE

LAURENT GOUZÈNES

GÉRARD ROUCAIROL (excusé)

PIERRE HAREN

#### IBM

ALAIN CHABRIER

ALEX FLEISCHER

### ➤ *Codes couleur:*

- Points importants soulignés par le GT

Les auteurs des questions sont indiqués par leurs initiales

## SYNTHÈSE

Un PoC a été développé à IBM pour illustrer des possibilités offertes par les outils d'optimisation dans le cadre du COVID. Il ne s'agit pas d'un outil opérationnel, mais bien d'un démonstrateur.

Nombre de difficultés rencontrées par les pouvoirs publics dans cette crise relèvent de la logistique (*supply chain*). Le PoC mentionné portait sur le transfert de patients lors du pic du mois de mars 2020. Les masques, les respirateurs, puis les tests, ont été d'autres exemples. **La campagne de vaccination à venir sera une autre étape délicate où il sera bon que les décisions soient prises en étant convenablement outillées.** Comme nous l'avons vu lors de plusieurs entretiens déjà, la modélisation épidémiologique n'est pas l'unique besoin en termes d'outillage informatique.

L'optimisation apporte des réponses utiles et effectives, mais elle demande un réel effort de modélisation : il faut identifier les leviers de contrôle, les critères qu'on veut optimiser, et les contraintes sous lesquelles l'optimisation doit se faire. Pondérer les divers critères est généralement un élément de choix politique.

Collecter les contraintes et les définir est la partie sur laquelle un effort significatif doit être fait au-delà du PoC réalisé pour en faire un véritable outil opérationnel. C'est une tâche de modélisation. **Comme dans tout développement algorithmique s'appuyant sur de la modélisation, c'est le coût de la modélisation qui prédomine, pour la conception et, plus encore, pour la maintenance (dans les cas où le modèle est susceptible d'évoluer). Une réflexion doit donc être menée sur la façon de rendre plus aisée ce travail de modélisation.** Nous développons les verrous correspondants à la deuxième section.

## DÉMONSTRATION DU PoC COVID-19 IBM

Cette démonstration a été faite pendant la réunion, commentée et discutée. On reproduit ici des renseignements extraits de pages et notebooks écrits par Florent Chabrier pour ce PoC. Il s'agit d'un PoC développé à IBM pour illustrer des possibilités offertes par les outils d'optimisation, pas d'un outil opérationnel dans le cadre de la pandémie de covid. On reviendra plus loin sur les éléments supplémentaires qui seraient à prévoir pour rendre un tel outil opérationnel. **Le texte anglais est un verbatim des pages Web de Florent Chabrier à l'adresse : . [Prototype Decision Optimization usage for Covid-19 decision making](#) (Mars 2020).** Les phrases en français sont rajoutées pour ce CR.

The problem of finding optimal ways to relocate sick people among areas can benefit from a combination of Machine Learning (ML) and Decision Optimization (DO). Based on the recent evolution of the number of critical reanimation cases in each area, **predictive models** can be trained to forecast the evolution per area on the coming days. This data, in addition to the capacity of the hospitals for each of the areas and some description of the constraints applying on the possible transfers can then be used in a **decision optimization** model. This schema where Machine Learning is used first to extract additional unknown information and Decision

Optimization is used to prescribe the best next actions is very common.

You can find [here an example notebook](#). This is a very simplistic prototype to illustrate how both technologies could collaborate, but a real solution would require a subject matter expert to ensure that the right data is used, and the right constraints and objectives are considered. This model formulation is very standard process in Decision Optimization projects and can take from days to months, depending on the complexity of the problem.

### Plan transfers between areas

This notebook shows how we can use Decision Optimization to optimize the move of affected people between hospitals to avoid being overcapacity.

The problem is solved per department.

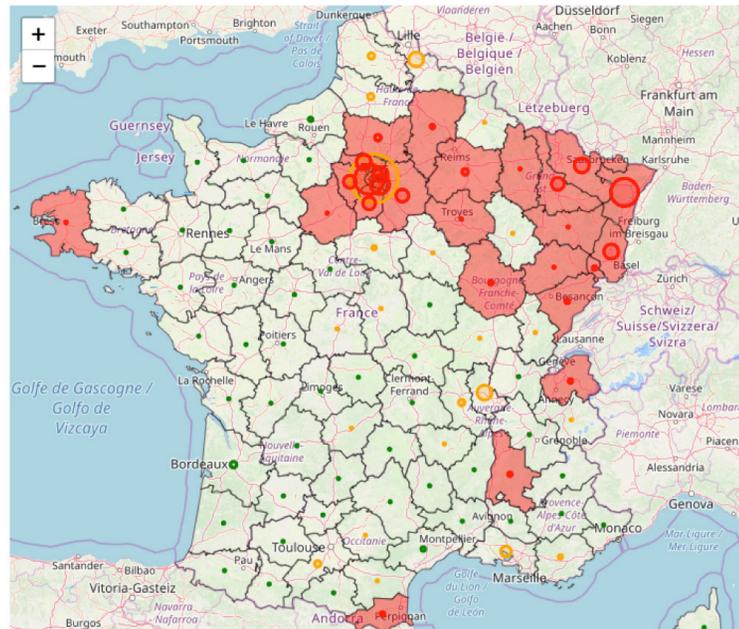
There are 5 parts in this notebooks:

1. Load the data form different places (departements, current situation, etc)
2. represent the current situation on a map
3. predict new cases to come for each department
4. plan transfers
5. display all the transfers from the solution

**DISCLAIMER:** this notebook is thought to demonstrate through an example how Machine Learning and Decision Optmization could be used, but it is partially based on fake data and knowledge of the real problem, and some significant work would be required to incorporate real data and real constraints and obejctives so that the outcome could be useful.

We use data from the French government [data.gouv.fr](https://www.data.gouv.fr/fr/datasets/donnees-des-urgences-hospitalieres-et-de-sos-medecins-relatives-a-lepidemie-de-covid-19/#) site: <https://www.data.gouv.fr/fr/datasets/donnees-des-urgences-hospitalieres-et-de-sos-medecins-relatives-a-lepidemie-de-covid-19/#> in addition to some imported data on the different French administrative “departments” (GPS coordinates of frontier and center).

Using Folium, we can easily represent this data on a map.



The circles show the number of reanimation cases in each department. Red circles show

above normal capacity. A very simplistic predictive model is trained for illustration. We use LinearRegression from sklearn, knowing obviously that the epidemy is not linear at all. But this gives an idea of how ML would work here.

Our hypothesis for the optimization model is that two different types of transfers can be done (ces catégories de transfert sont définies à l'aide de paramètres dont le choix et la valeur appelleraient plus d'élaboration):

- long distance transfers (planes, trains) with the number of transfers limited over the whole country, several people can be transferred at a time ;
- short distance transfers (ambulances) with the number of transfers limited per area, and with just one person at a time.

The decision variables are the transfer links to be used in the best (optimal) transfer plan (an integer variable for the number of persons transferred and a binary variable indicating whether the link is used or not, with a constraint linking both).

Constraints are then formulated (we show a sample here):

```
# Initial state
mdl.add_constraints(occupancy_vars[d, 0] == initial[d] for d in deps)

# structural constraint between user_link and link
mdl.add_constraints(use_link_vars[d, d1, t] == (link_vars[d, d1, t]
>= 1) for d in deps for d1 in deps for t in transfer_periods)

# Short transfers bounds
mdl.add_constraints(link_vars[d1, d2, t] <= 1 for d1 in deps for d2
in deps if not is_long[d1][d2] for t in transfer_periods)

# number of transfers from a department less than current number of
cases
mdl.add_constraints(mdl.sum(link_vars[d, d1, t] for d1 in deps) <=
occupancy_vars[d, t] for d in deps for t in transfer_periods)

# maximum number of LONG transfers
mdl.add_constraints(mdl.sum(use_link_vars[d1, d2, t] for d1 in deps
for d2 in deps if is_long[d1][d2]) <=
MAX_NB_LONG_TRANSFERS_PER_PERIOD for t in transfer_periods)

# maximum number of SHORT transfers
mdl.add_constraints(mdl.sum(use_link_vars[d1, d2, t] for d1 in deps
if not is_long[d1][d2] for t in transfer_periods) <=
MAX_NB_SHORT_TRANSFERS_PER_DEPARTMENT for d2 in deps )

# conservation constraints including new cases to come
mdl.add_constraints(occupancy_vars[d, t+1] == new_rea[d][t] +
occupancy_vars[d, t] + mdl.sum(link_vars[d1, d, t] for d1 in deps) -
mdl.sum(link_vars[d, d1, t] for d1 in deps) for d in deps for t in
transfer_periods)
```

Ces contraintes sont exprimées dans le langage python en utilisant l'api docplex de l'outil d'optimisation (CPLEX). Ces contraintes représentent des règles (sanitaires, logistiques...) qui doivent être respectées par toute politique de transfert.

Main objective :

```

final_overcapacity = mdl.sum(mdl.max(0, occupancy_vars[d,
NB_PERIODS] - capacity[d]) for d in deps)
mdl.add_kpi(final_overcapacity)

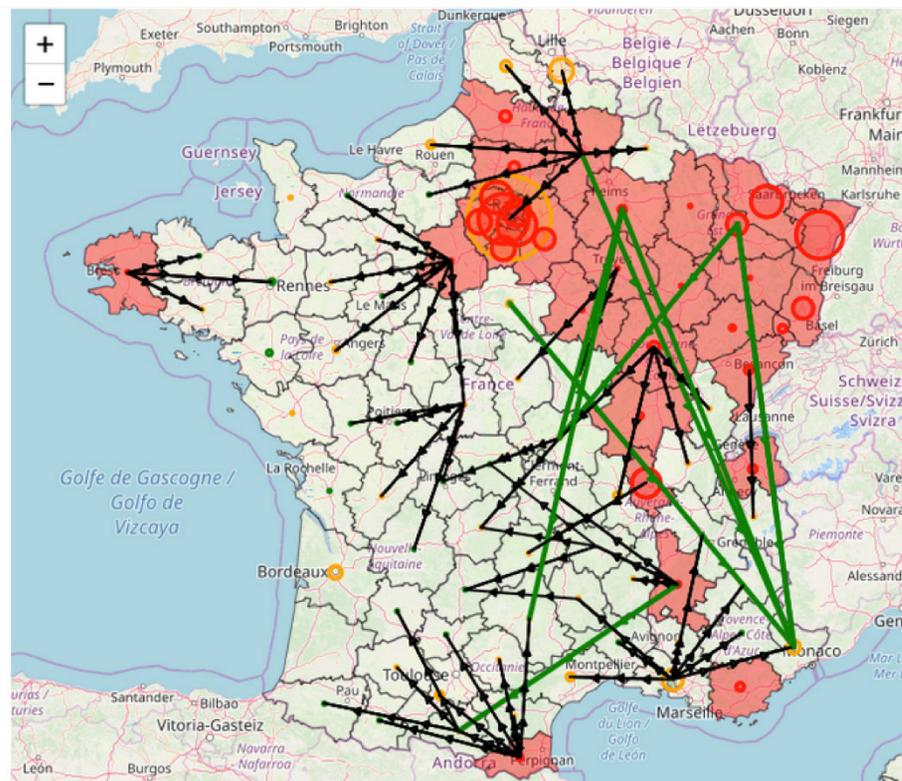
nb_long_transfers = mdl.sum(use_link_vars[d1, d2, t] for d1 in deps
for d2 in deps if is_long[d1][d2] for t in transfer_periods)
mdl.add_kpi(nb_long_transfers)

nb_short_transfers = mdl.sum(use_link_vars[d1, d2, t] for d1 in deps
for d2 in deps if not is_long[d1][d2] for t in transfer_periods)
mdl.add_kpi(nb_short_transfers)

mdl.minimize(1000 * final_overcapacity + 10 * nb_long_transfers +
nb_short_transfers)
    
```

The main objective is to reduce the total overcapacity on the areas, but we should also limit unnecessary transfers (again, this would need further elaboration):

Using the same Folium package, we can represent the proposed transfers on a map. The long-distance transfers are represented in green and the short distance ones in black.

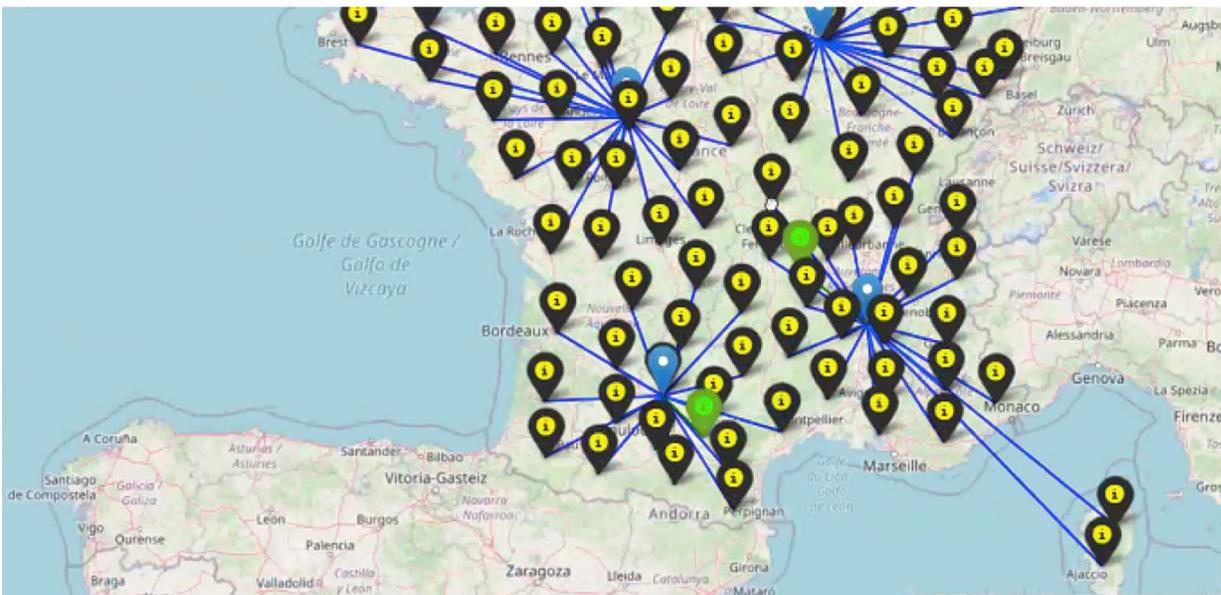


Dans la suite de ce CR on regroupe des éléments de discussion et une réflexion sur les efforts supplémentaires qui seraient requis pour construire une vraie solution pour ce type de problème de décision.

## DISCUSSION

Il existe bien d'autres problèmes du même type dans la crise du COVID

Ceci était un cas d'utilisation sur le transfert des malades, question d'actualité en mars 2020. Depuis, on en a traité le cas des masques, avec une réutilisation d'à peu près la même démarche.



En fait, ces deux exemples sont représentatifs de problèmes de *supply chain* (chaîne logistique). La crise de la covid en présente d'autres instances (les respirateurs, les tests...), et nous reviendrons plus loin sur le cas d'actualité des vaccins (fin 2020).

### QUELS SONT LES VEROUS POUR FAIRE UN VRAI OUTIL OPÉRATIONNEL ?

#### ➤ Les contraintes constituent l'essentiel de l'effort de modélisation

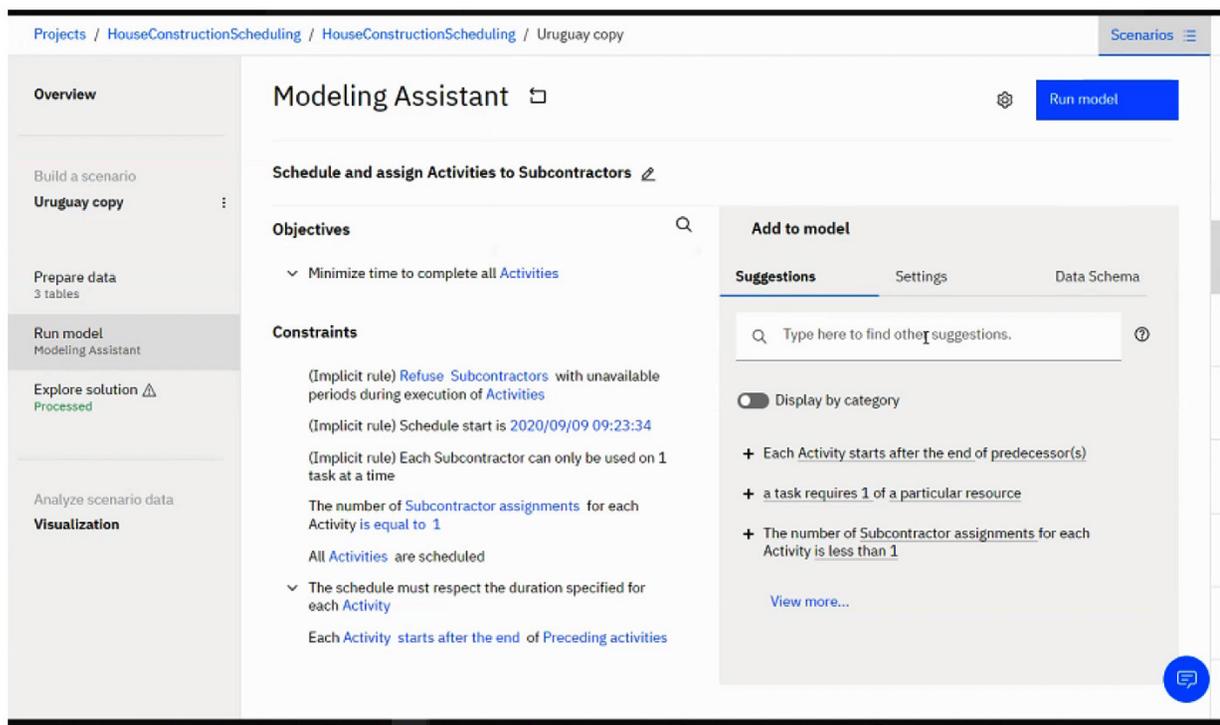
L'outil d'optimisation demande de définir les contrôles, les critères à optimiser, mais aussi les **contraintes**. C'est ce dernier point qui semble le plus difficile. **Il faut trouver les bonnes règles, les bonnes contraintes. Et ça c'est de la modélisation relevant du domaine de la logistique. Cet effort de modélisation est requis, et la vraie difficulté est de le faire, pas de montrer que c'est faisable.** L'effort de réalisation de tels modèles se compte en jours, grâce à la technologie et à ses progrès. La programmation n'est plus un problème, sauf pour ce qui concerne la facilitation de l'expression des contraintes. Mais collecter l'expertise et trouver les bonnes règles reste coûteux et pénible.

Comme toute méthode de conception utilisant de la modélisation, c'est rapidement le coût de la modélisation qui devient prédominant au regard de l'effort d'ensemble. Ceci handicape le recours aux modèles (et fait, en retour, le bonheur des approches s'appuyant uniquement sur les données). Une conception utilisant des modèles demande cet effort de modélisation. Il y a plus ennuyeux : dans les cas où ces modèles demandent à être mis à jour, l'effort de maintenance peut devenir rédhibitoire. Un effort particulier doit donc être consacré à faciliter la tâche de modélisation.

Malheureusement, les caractéristiques de la crise, et surtout son côté inédit, font que la collecte de connaissances par fouille de données sur une base de documents contenant la connaissance (de manière non structurée), ne semble guère faisable compter sur l'expertise semble plus réaliste. Cependant, dans le cas présent des contraintes, les angles d'attaque suivants existent :

- les règles peuvent être rentrées incrémentalement ; le travail de modélisation n'est donc pas à refaire en totalité à chaque mise à jour ;
- les règles sont à obtenir de la part d'experts pour chacun des divers aspects du problème d'optimisation. Ces experts ne seront pas, en général, des experts du langage d'expression des contraintes. Il faut donc prévoir des interfaces constituées de langages métier pour exprimer les contraintes dans les différents aspects.

➤ *Les règles exprimées en langage naturel :*



La copie d'écran ci-dessus représente en colonne du milieu des règles en langage naturel exprimées par des experts. L'outil possède une base de *patterns* (patrons de règles prédéfinis)

parmi lesquels il propose en retour un plus proche voisin (selon une distance du domaine de la fouille de texte), qui peut être validé ou non par l'expert.

Plusieurs APIs sont possibles pour rentrer les contraintes et objectifs. Cela peut être fait en python comme dans l'exemple PoC. Il faut alors expliquer aux médecins et autorités ce qu'il y a derrière le code. Cela peut être fait en OPL (langage haut niveau de modélisation au-dessus de CPLEX et dans ce cas les médecins/autorités peuvent presque être autonomes dans la lecture et le commentaire voire même dans la critique du modèle. On peut aussi utiliser le Modeling Assistant qui permet de partir d'un modèle générique et de l'adapter à un domaine particulier à l'aide d'interaction en langage naturel complétée par une interprétation intelligente des noms des champs dans les données.

### ➤ *Collecte de l'expertise*

**D**eux grandes familles d'approches peuvent être envisagées et combinées.

#### Collecte dans des bases d'information relevant de la logistique

**D**ans les problèmes de décision logistique, il existe nombre de questions qui sont de nature assez générique, et peuvent donc être importées de domaines plus connus. Se posent alors les difficultés habituelles : si plusieurs domaines sont concernés, les connaissances peuvent être exprimées sous forme hétérogène, structurées selon des ontologies différentes, etc. Ceci ne constitue pas en soi une difficulté nouvelle.

#### Travail collectif d'experts en « *crowd sourcing* » dans un réseau social

**C**ette approche semble particulièrement appropriée pour une crise inédite telle que le COVID. Le sentiment de l'urgence peut faciliter la mobilisation d'un groupe d'experts de diverses spécialités qui concourent à définir collectivement les règles. Pour cela, il faut disposer des éléments suivants :

- un support facilitant l'expression de règles ou contraintes dans le mode de pensée des divers domaines, voir plus haut ;
- un réseau social privé qui sert de support à l'interaction collective. Ce support reste à définir et réaliser – la technologie existe.

### ➤ *Une certaine méconnaissance de ce que l'optimisation peut apporter*

**C**ontrairement au diptyque {intelligence artificielle, données} qui bénéficie d'une large popularité dans les médias, les milieux politiques et même le public, le domaine de l'optimisation reste largement un domaine de spécialistes et ses possibilités restent méconnues.<sup>93</sup> Ceci est une vraie difficulté qui, d'ailleurs, avait motivé le développement du PoC COVID à IBM.

93 L'optimisation n'est pas le seul domaine dans cette situation. C'est plutôt une situation assez répandue.

### **LA CAMPAGNE DE VACCINATION : LA PROCHAINE DIFFICULTÉ ?**

Il est clair que la campagne de vaccination qui s'annonce présente une sérieuse combinaison de chausse-trappes potentielles:

- le pouvoir politique doit définir dans quel ordre les différentes sous-populations se verront proposer la vaccination (au 24/11/2020 on a appris que les plus de 65 ans seraient considérés en premier, et non pas les soignants comme il avait été suggéré) ;
- il aura la responsabilité de définir quels vaccins seront déployés, et quand ;
- la logistique correspondante devra suivre : information des populations, préparation des locaux, disponibilité des vaccins...

Il paraît absolument critique que les pouvoirs publics puissent disposer d'une argumentation bien étayée sur la base de quoi de telles décisions seront prises. Un terrain d'intervention désigné pour la problématique abordée dans cet entretien.

Un expert sur les vaccins est Pascal Brandys, IPC, ancien PDG fondateur de Genset, qui vit maintenant à San Diego (Ca) et qui a créé Phylex BioSciences ([www.phylexbio.com](http://www.phylexbio.com)) qui est une start-up avec un brevet sur le seul vaccin qui soit résistant aux mutations standard du virus covid. Pascal pourrait être contacté pour décrire les paramètres nécessaires à la compréhension du problème du séquençage de la vaccination dans le monde (quel vaccin, quels sous-groupes de population, quels risques, quels problèmes logistiques, etc.)

Si nous recommandons au gouvernement de prendre une approche « d'optimisation explicable » à ce problème, elle aura l'avantage de mettre sur la table la liste des contraintes auxquelles nous sommes soumis, ce qui sera nécessaire dans un débat public qui apportera sans nul doute son lot de théories complotistes.



## ANNEXE 5 — ENTRETIEN AVEC JEAN-MICHEL LASRY — *MEAN FIELD GAMES* (MFG)

*(26 janvier 2021)*

*Notes d'Albert Benveniste et Laurent Gouzènes, revues par Jean-Michel Lasry.*

### ➤ *Participants*

ALBERT BENVENISTE

LAURENT GOUZÈNES

JEAN-MICHEL LASRY

GÉRARD ROUCAIROL

### ➤ *Codes :*

- Les questions sont indiquées en italique et les réponses en roman
- Les auteurs des questions sont indiqués par leurs initiales

## SYNTHÈSE ET MESSAGE DE JEAN-MICHEL LASRY

Quand on parle de modélisation épidémiologique, il s'agit de modéliser l'interaction entre les politiques sanitaires, les comportements socio-économiques et la maladie. Quand on parle de modèle d'épidémiologie, et de la nécessité de prendre en compte la dimension humaine des agents,

- c'est parce que l'on pense non pas à la maladie seule ;
- mais à l'interaction de la maladie avec les comportements sociaux, économiques, et avec les politiques sanitaires ;
- et aussi à l'interaction avec les stratégies des producteurs de vaccins et traitements.

Autrement dit, l'enjeu de modélisation pour les épidémies (pandémies) humaines (surtout si la modélisation doit servir à prendre des décisions de politique sanitaire) c'est d'incorporer à l'intérieur du modèle tout le contexte socio-économique et réglementaire, et non pas de se contenter de faire un modèle qui serait valable pour une population d'animaux sauvages.

Pour le répéter encore, les politiques sanitaires (et leur acceptabilité pour le corps social, et les réactions à ces politiques selon les cultures) doivent être un élément central de n'importe quel éventuel modèle covid-19. Comme ces modèles ne sont pas faits, les pouvoirs politiques prennent actuellement leurs décisions en suivant leurs intuitions et quelques évaluations pifométriques des conséquences de leurs décisions.

Pour finir, et pour élargir le débat, voici une expérience de pensée sur l'interaction entre la maladie et l'intérêt des producteurs de vaccins. Si l'on réfléchit à modéliser la stratégie égoïste des producteurs de vaccins, on s'aperçoit que la stratégie égoïste optimale (celle qui maximise le profit espéré actualisé) serait de ne pas produire trop de vaccins de sorte que le virus ait le temps de muter, ce qui renouvelle la clientèle pour de nouveaux vaccins (c'est l'équilibre auquel on est arrivé pour la grippe – par contre la poliomyélite et d'autres maladies ont été éradiquées ; pour comprendre cette différence, il serait sans doute possible de faire une modélisation et de vérifier que selon les paramètres, l'équilibre est différent). Heureusement dans le cas de la covid-19, il y a une double sécurité :

- d'une part il semble que les laboratoires ne puissent pas se permettre d'être égoïstes ;
- et par ailleurs la concurrence les prive de cette option stratégique.

Cependant la mollesse des gouvernements (seuls clients acceptés actuellement par les laboratoires) aboutit à une solution qui semble proche de l'optimum des producteurs égoïstes.

## IMPORTANCE DES ANTICIPATIONS DANS LE COMPORTEMENT D'UNE POPULATION

Parabole du « pas de route<sup>94</sup> » elle illustre le comportement des foules pour le parallèle avec la covid, on fait le confinement au moment où les gens commencent à le demander.

Je vois qu'il y a une grande tradition de modélisation et de simulation dans le domaine de l'épidémiologie. 98 % des modèles développés s'appliquent aussi bien aux humains qu'aux animaux ou végétaux. On n'introduit aucune modification résultant du caractère humain et social (sinon via le calage de paramètres sur des données). Or, le fait d'avoir affaire à des humains change les choses. Quand un gouvernement annonce une décision, la population s'adapte à la décision (et fait éventuellement un peu autre chose). En économie on est habitué à ça depuis longtemps. Cette ignorance des anticipations est donc un gros défaut des modèles épidémiologiques. Le calage des paramètres compense en partie ces défauts sur le court-terme, mais ça ne vaut plus pour le *what-if* à plus long terme. *What-if* : que se passe-t-il si j'annonce telle ou telle de mesure de confinement ? quel est le comportement en retour des humains ? Il y a donc une erreur méthodologique fondamentale dans les approches actuelles.

Il y a 3-4 ans, des chercheurs (Turinici et al<sup>95</sup>) qui travaillent dans la communauté MFG ont fait un modèle sur la vaccination et le comportement des individus vis-à-vis de la vaccination<sup>96</sup>. À cette époque l'attention n'était pas forte et c'est passé un peu inaperçu. En mars 2020, un chercheur a fait un modèle d'équilibre (équilibre de Nash au sens de la théorie des jeux, utilisé pour expliquer les boucles décisions-prix qui se produisent en économie). En épidémie, la situation est similaire. Les populations sont aussi à un équilibre, mais avec des quantités macroscopiques qui ne sont pas des prix. Un exemple de quantité est l'appétence à prendre le métro en fonction de la densité de remplissage perçue. Beaucoup d'équilibres collectifs sont ainsi déterminés collectivement.

Il résultera de la pandémie de la covid-19 des impacts psychologiques de long terme qui vont s'incruster (procédures de sécurisation). Prenez l'exemple des bureaux à la Défense. Si le gouvernement annonce que les bureaux à la Défense c'est fini à cause du télétravail, la valeur des commerces va chuter, le quartier va devenir sinistré, en conséquence de quoi les cadres ne voudront pas y revenir au total la prédiction se réalisera. Si le gouvernement prédit l'inverse, ça peut faire revenir les cadres. Donc il y aurait deux équilibres de Nash selon les effets d'annonce.

C'est ce que nous cherchons à mettre en évidence avec les champs moyens. Pour citer un autre exemple, sur la route, ou bien tout le monde circule à gauche ou bien tout le monde circule à droite : il n'existe que deux équilibres qui s'installent spontanément, même en l'absence de code de la route. Quand il y a un grand choc on peut changer d'équilibre. C'est ce changement d'équilibre sous l'effet de décisions annoncées que l'on peut chercher à modéliser. Ce type de grand changement ne peut pas être prédit avec un modèle ne prenant pas en compte

94 Quand la troupe est fatiguée et commence à donner des signes de désorganisation, le commandement ordonne le « pas de route », plus relax...

95 Voir <https://hal.archives-ouvertes.fr/hal-02545930v2/document> Contact rate epidemic control of COVID-19: an equilibrium view. Romuald Elie, Emma Hubert, Gabriel Turinici.

96 Si je vois que les autres se vaccinent, je pense que je n'ai plus besoin de me vacciner, ce qui diminue le nombre de candidats au vaccin.

la « physique des foules », même si l'on en compense les imperfections par un recalage des paramètres sur des données.

*LG : le fait qu'il y ait une dispersion sur le territoire en 2D me fait penser au modèle d'Ising en physique ?* Oui. Dans notre communauté, on se pose aussi la question des interactions dans des espaces plus complexes comme des graphes (y compris les réseaux sociaux). Les techniques d'étude sont délicates et récentes. On regarde ça avec Pierre-Louis Lions depuis deux ans et on progresse. C'est une question difficile.

Ça peut avoir un effet de casser les routines dans les tâches de modélisation (cf. le SIR) Demongeot et d'autres ont territorialisé le SIR c'est étonnant qu'on ait attendu si longtemps avant de regarder ces extensions. C'est vrai à l'échelon international.

Sur un plan plus technique, dans un modèle MFG la situation typique est que dans un tel modèle on a des équations différentielles qui traduisent les dynamiques réelles mais il y a en plus une formulation des dynamiques futures anticipées par la communauté. Donc à chaque instant le monde se déplace en fonction du passé et de l'anticipation du futur par la communauté. Chacun choisit sa décision en fonction de ça. Ça se traduit par deux équations :

- Forward qui décrit la trajectoire réelle ;
- Backward pour la traduction des anticipations.

Ces deux paquets d'équations sont couplés.

*AB : à quelle distance est-on de faire une utilisation en grand sur les vrais problèmes ?*

En fait quand on est membre d'une foule on n'a aucune influence directe sur la foule. Il n'y a pas d'explosion combinatoire. Tout est simplifié par la moyennisation. Il faut faire des simplifications et ne pas se tromper sur les simplifications. Modéliser est un art. Après c'est comme tous les modèles. Si on fait des compartiments, on finit par avoir plusieurs dizaines de populations et il va falloir beaucoup de données.

## UTILISATION PRATIQUE DE L'APPROCHE MFG ?

Toutes ces techniques sont compliquées. Je ne suis pas sûr qu'il y ait un usage immédiat. Mais je pense que les MFG devraient apporter la dimension humaine aux prédictions *what-if*. C'est comme l'économie avec la boucle prix-comportements c'est plus compliqué que d'exercer un contrôle sur un système physique. Cf. la pénurie du sucre dans les années 1980, largement amplifiée par les effets d'annonce. L'effet d'annonce est énorme. Cette importance a fini par être comprise en économie. Mais la même chose vaut ailleurs et ça n'est pas admis ailleurs, y compris pour l'épidémiologie.

Je trouve intéressant ce que Dassault Systèmes fait avec sa plate-forme 3DEXPERIENCE. Mais je pense que, pour les modélisations multi-agents qu'ils peuvent y développer, ils doivent avoir un problème avec la dimension humaine. Dans les modèles de foules, des modèles de particules en interaction sont utilisés, mais sans prendre en compte les anticipations. Je sais qu'il existe des essais de modélisation sur l'évacuation des foules à Dubaï, qui prennent en compte les anticipations. On aboutit à un équilibre de Nash.

Attention, l'équilibre de Nash est intéressant parce qu'il explique ce vers quoi converge la communauté. En revanche, cet état de convergence peut être un désastre pour la communauté (cf. le dilemme du prisonnier<sup>97</sup>). L'équilibre de Nash est parfois loin de ce qu'un planificateur voudrait comme résultat. C'est en restreignant l'espace de contraintes des gens que l'on améliore l'équilibre de Nash ça se voit très bien dans la modélisation des mouvements de foule<sup>98</sup>.

Donc c'est important de prendre en compte les comportements sociaux ce qui justifie la considération des MFG.

On peut avoir envie de rajouter des règles de façon à rendre optimal l'équilibre de Nash. Le rôle du régulateur est de dessiner les règles de façon à ce que l'équilibre qui en résulte soit le meilleur possible pour la communauté. Ce rôle de régulation est donc important. Les règles améliorent l'équilibre global. Il est donc important d'avoir une bonne crédibilité de la parole publique (pour que les règles soient suivies d'effet).

### DISPONIBILITÉ DES DONNÉES

On a la chance en France d'avoir une seule compagnie d'assurance maladie (la sécurité sociale) qui a toutes les données donc on pourrait tout explorer, en principe, modulo le respect de la protection de la vie privée. On a tous les actes médicaux qui sont archivés par la sécu (fichier SNIRAM). Les États-Unis n'ont absolument pas ça. Mais ces données de la sécu sont gelées et dorment totalement sous toutes sortes de strates et de passivités.

*GR : Bacri s'est heurté à une difficulté de fond sur ce sujet.* Mais il ne convenait pas pour cette fonction. Sur le papier (son CV) il est l'homme de la situation. Mais il a montré une forme de procrastination, il n'est pas du genre à hurler contre le scandale que constitue le gel de ces données. Pas suffisamment, en fait.

Sur la publication de ces données, c'est la sécu qui freine des quatre fers. Comme souvent, les données sont une question de lead et de pouvoir. Dans notre groupe de travail on va concentrer nos feux là-dessus. Disposer de ces données permettrait de faire des progrès gigantesques. Pour de mauvaises raisons ça semble impossible. Le [health-data hub](#) aurait pu être la solution, mais le résultat est que ça fonctionne comme une défense en attendant il ne se passe rien. C'est criminel de laisser dormir toutes ces données. On pourrait faire des milliers d'études qui pourraient sauver des vies. Le verrou le plus absolu c'est ça.

*GR : il en existe un ouvert du côté de Nantes.*

La mise sur la table des modèles MFG exige de telles données. Pour avoir de vrais modèles MFG il faudrait vraiment ces données. Si je n'ai pas de données je ne peux pas prouver que

97 Le dilemme du prisonnier caractérise en théorie des jeux une situation où deux joueurs auraient intérêt à coopérer, mais où, en l'absence de communication entre les deux joueurs, chacun choisira de trahir l'autre si le jeu n'est joué qu'une fois. La raison est que si l'un coopère et que l'autre trahit, le coopérateur est fortement pénalisé. Pourtant, si les deux joueurs trahissent, le résultat leur est moins favorable que si les deux avaient choisi de coopérer. Source : [https://fr.wikipedia.org/wiki/Dilemme\\_du\\_prisonnier](https://fr.wikipedia.org/wiki/Dilemme_du_prisonnier)

98 Dans une ville aux États-Unis ils avaient mis sur internet les temps d'attente aux urgences, ce qui a eu un effet régulateur dans ce cas il n'y a pas d'effet pervers. Sur les hôpitaux, c'est différent les gens se disaient : s'il y a une longue attente c'est que l'hôpital est bon donc j'y vais, d'où un effet indésirable.

j'ai raison car la vérité terrain manque pour comparer. Si je commence à faire un modèle basé sur la peur de la maladie, il faudra à la fin que je puisse le confronter à des données riches.

*LG : par contre, si je reprends l'exemple des eaux usées, on n'a pas besoin de grandes masses car ces données sont déjà agrégées. Donc tout dépend de ce qu'on veut avoir comme données.*

Il est vrai qu'en plus des données de la sécu, il y a les données eaux usées. Mais aussi les données Google-trend (mots sur lesquels on pose des questions). Au début de 2020 le mot « toux » montrait une avance de 10-15 jours sur ce qui était annoncé par les hôpitaux. Cf. [Google-flu-trends](#), mais les gens se sont comportés de manière tordue sur ce point, questionnant avec certains mots-clés pour avoir des réponses sur autre chose, de sorte que le service a été stoppé. À partir de juin, le mot « toux » a cessé d'être important. Le mot « toux » a typiquement un plus bas net fin août et rebondit début septembre ceci ne peut s'expliquer que par la rentrée des classes. Énormément de données comme ça sont disponibles.

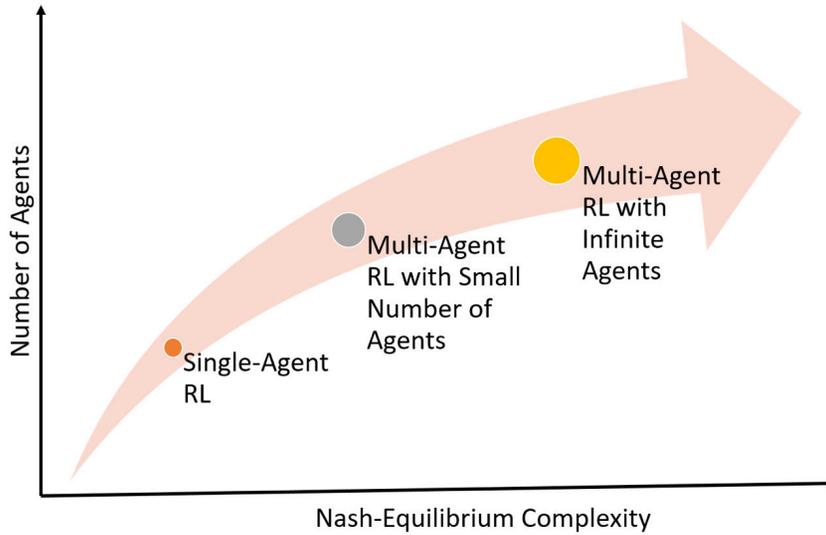
« Toux » est aussi un bon indicateur de l'angoisse. « Toux » a un pic entre 1h et 5h du matin, comme les mots « infarctus » et « mort ». On pourrait mettre dans les modèles d'épidémiologie les angoisses des gens.

Pour comparer la situation avec les données de la sécu, les données Google/Amazon sont disponibles, il suffit de les acheter (elles ne sont pas occultées pour raison administratives obscures). Un autre exemple est l'ensemble des courbes Google des déplacements. On a des données équivalentes de la part de opérateurs de télécommunications en France (on en a parlé au printemps). Mais les chercheurs n'avaient pas d'argent pour acheter ces données et donc les opérateurs ont été découragés de mettre ces données à disposition. En revanche Google etc., donnent gratuitement ce genre d'information comme élément de soft power. Je ne sais pas pourquoi Orange ne fait pas cadeau de ces données.

#### *QUELQUES DÉTAILS SUR LES TRAVAUX TURINICI ET AL.*

**E**n préambule, on peut rapprocher cette nouvelle approche d'autres approches dites multi-agents. Dans ces approches multi-agents, on modélise à la fois la propagation du virus et les comportements individuels en réaction au virus. On a donc un modèle avec une propagation de virus et des agents.

Ceci est schématisé dans la figure suivante :



<https://hal.archives-ouvertes.fr/hal-02545930v2/document> Romuald Elie, Emma Hubert, Gabriel Turinici .

On part du modèle SIR usuel tel que décrit dans le schéma ci-dessous,



$$\begin{cases} dS_t = -\bar{\beta}_t S_t I_t dt \\ dI_t = \bar{\beta}_t S_t I_t dt - \gamma I_t dt \\ dR_t = \gamma I_t dt \end{cases}$$

à ceci près que le facteur de propagation  $\bar{\beta}$  n'est plus fixé *a priori* comme étant un paramètre du modèle, mais est le résultat du comportement moyen de la population. La description de ce comportement moyen est précisément la contribution originale du modèle MFG. Il résulte d'un équilibre de type Nash pour une grande population (supposée homogène). Plus précisément, si  $P_t^\beta = P(\tau \leq t)$  est la probabilité que l'instant d'infection  $t$  soit inférieur ou égal à  $t$ , on a  $dP_t^\beta = \beta_t I_t (1 - P_t^\beta) dt$  pour  $0 \leq t \leq T$ . On introduit alors le coût pour un individu de la population du choix de son coefficient  $\beta$  personnel :  $C(\beta, \tau) = \int_0^{\inf(\tau, T)} c(\beta_s) ds + r_I 1_{\tau \leq T}$  où la distribution de  $\tau$  est définie par la dynamique de  $P_t^\beta$  et donc dépend indirectement de  $\bar{\beta}$ . Chaque individu cherche à minimiser par rapport à son coefficient individuel  $\beta$  le critère suivant, en prenant en compte la population moyenne :  $\bar{C}(\beta, \bar{\beta}) := E(C(\beta, \tau))$ . L'équilibre atteint est de Nash, mais pour une population de taille « infiniment grande » (il y a donc un passage à la limite).

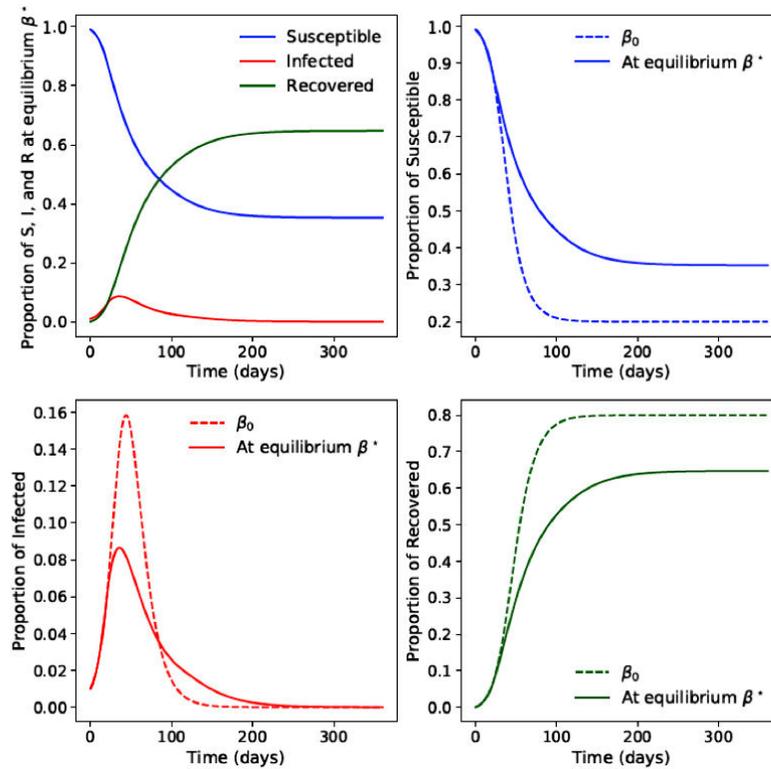


Figure 3: Evolution (in proportion) of Susceptible, Infected and Recover classes, using the parameters described in Table 1. Solid lines represent the evolution at the Mean Field Nash equilibrium, *i.e.*, with transmission rate  $\beta^*$ . The evolution at equilibrium is compared to the epidemic dynamics with constant transmission rate  $\beta_0$  (dashed lines).

Les courbes ci-dessus permettent de comparer les résultats du modèle SIR classique (trait pointillé) avec ceux de ce modèle MFG (trait continu). On voit donc bien la différence.

### QUELLES RECOMMANDATIONS DEVRAIT-ON FORMULER ?

➤ *Il faut modéliser*

La modélisation est importante pour faire des *what-if*. Les projections statistiques ne permettent pas de deviner les nouveaux équilibres. La modélisation est importante mais est un art délicat. Je vous enverrai par mail quelques noms. Il existe quelques experts qui savent faire des modèles de jeux à champ moyen mais qui doivent aussi apprendre les démarches des modèles classiques. Inversement il faut que les modélisateurs classiques aient conscience des trous dans la raquette dans leur approche. Il faut donc mettre les communautés en rapport.

**L'ACCÈS AUX DONNÉES EST INCONTOURNABLE POUR VALIDER ET CONFRONTER  
LES MODÈLES.**

**L**a grande chance qu'on a est que c'est rare d'avoir des chocs et c'est dans les chocs qu'on voit si les modèles sont bons.

➤ **Sur les passerelles entre communautés existantes ?**

**R**ien pour l'instant. Sur les MFG, la passerelle qui existe est avec l'économie. Avec cette communauté-là, on a eu, à la fois, de la réticence et de l'écoute. Sur le plan conceptuel, ça ne leur semblait pas nouveau, mais ils ont ensuite changé d'opinion : les concepts ont finalement été admis. Il reste une vraie réticence, due à la difficulté mathématique.

Dans d'autres communautés ça a été plus difficile. Ça n'était pas la peine d'y penser dans l'épidémiologie avant la présente crise (comme expliqué plus haut, on a échoué sur les modèles de vaccination). On n'a pas essayé de contacter l'Inserm. On n'a fait des efforts qu'avec les communautés qui venaient à nous (en France et aux États-Unis). J'ai fait avec quelques personnes un modèle sur l'industrie pétrolière mondiale qui a très bien fonctionné en 2020. Mais c'est du macro-économique pluriannuel donc ils ne se sont pas jetés là-dessus aussi, comme c'est difficile techniquement, on n'a pas eu un succès fou. On a été agréablement surpris par le monde des télécommunications : les objets connectés manipulés par des humains réagissent comme les humains de ce point de vue.



## ANNEXE 6 — ENTRETIEN AVEC PATRICK GUÉRIN, OPENHEALTH : UNITÉ COMETE : ANALYSE DES EAUX USÉES

(11 mars 2021)

Notes de Albert Benveniste et Laurent Gouzènes, Revisés par

### ➤ **Participants**

#### **Académie**

**LAURENT GOUZÈNES**

**ALBERT BENVENISTE**

#### **OpenHealth**

**PATRICK GUÉRIN**, vétérinaire de formation collecte de données de santé, dossiers médicaux (humains)

### ➤ **Codes:**

- Les questions sont indiquées en italique et les réponses en roman
- Les auteurs des questions sont indiqués par leurs initiales

## OPENHEALTH

Openhealth<sup>99</sup> est une société privée basée à Vannes. Nous vendons des données de consommation médicamenteuse. On suit, par exemple, les allergies, le pollen et on affiche les évolutions sur une carte de France. En plus nous avons des partenariats de recherche, dont COMETE<sup>100</sup>.

Nous avons opéré les deux plates-formes mondiales de grippe sévère dans le monde les données arrivaient sur notre serveur nous avons pour mission de collecter les données et de les restandardiser pour alimenter le réseau de santé. Ça sert à l'OMS pour définir la composition du vaccin contre la grippe par exemple.

## RÉSEAU COMETE

### ➤ Les fondateurs



- largest of the five elite units of the French civil security forces
- specialized in Nuclear - Radiological - Bacteriological - Chemical - Explosive risks (NRBCE)
- >18'000 wastewater samples analysis during the Covid-19 outbreak
- expertise in sampling, filtration and extraction
- technical analysis capabilities (including variant screening)
- specialized in aggregation, processing and literacy of **health data operating data monitoring for French Health Authorities** (HAS, ANSM) and international surveillance networks (GIHSN)
- **coordination and development of the network**,
- **real-time data platform**, hosted in Vannes (FR) on a highly secured site
- compliance with the French Data Protection Authority

### A network focused on operational excellence – Associate members

**BIOMERIEUX** (Systems and reagents for first-line SARS-CoV-2 screening)  
**C4DX** (Overflow analysis)

**BIOSELLAL** (Variants screening kits),  
**SICPA** (data process integrity)

### Some operations

**N**ous avons fait différentes études pour la covid. On est spécialistes de *literacy* de la donnée. Nous avons travaillé sur les mesures qui sortaient des tests, les résultats étaient très hétérogènes. Nous sommes intervenus rapidement pour stocker les données des mesures de contamination dans l'eau. Nous avons ainsi créé avec le BNPM (Bataillon des marins pompiers de Marseille) le réseau COMETE. Ils disposent d'une des unités d'élite de la sécurité civile spécialisée dans les risques nucléaires, radiologiques, biologiques, et chimiques. Après de nombreux essais, ils ont mis au point une méthode de détection de virus dans les eaux usées. Les pompiers étaient intéressés à guider rapidement des décisions de nature logistique et opérationnelle (où mettre des moyens), par une prévision épidémiologique court terme.



////////////////////

99 <https://www.openhealth.fr/>

100 <https://infodujour.fr/societe/41090-marseille-lunite-COMETE-ou-lefficacite-contre-la-covid-par-la-prevention>

➤ **Exemples**



**1- Cluster avoided in a nursing home in Moselle, France, within 48 hours.**

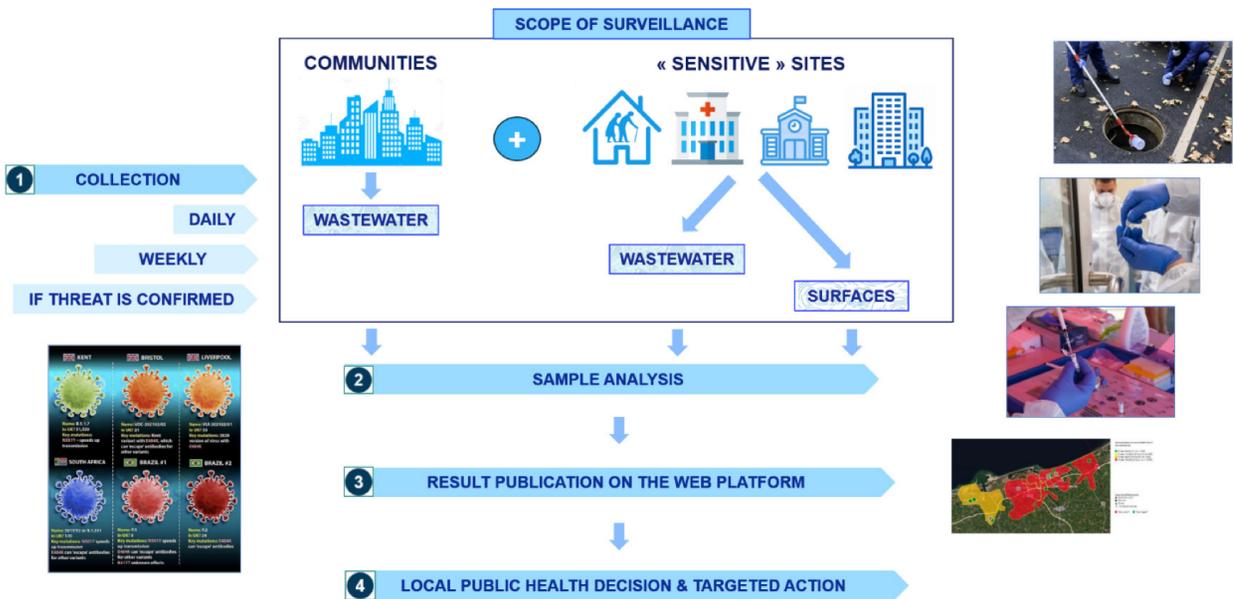
Day 1 AM : rescue intervention of French BMPM in high risk cities of the North-Est region, immediate sampling in wastewater plants for each district of the community.  
 Day 1 PM : analysis and mapping of districts where the variants were circulating the most  
 Day 2 AM : sampling of wastewater at sensitive sites in positive districts  
 Day 2 AM : identification of contaminated sensitive sites, surface testing of areas inside the nursing home  
 Day 2 PM : individual PCR testing of the population living or working in the contaminated areas.  
 Day 2 PM : isolation of the unique individual who was positive to COVID-19, asymptomatic and super-spreader. Decontamination of the nursing home. Back to normal operation. Cluster avoided.

**2- With the same method, a cluster was also avoided in a nursery in Marseille, France, and children were able to go back to school safely after 72 hours.**

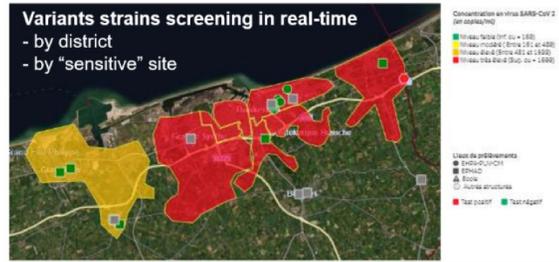


Les pompiers de COMETE mesurent la concentration dans certains locaux et peuvent ainsi identifier des sources assez localisées (jusqu'à un étage d'un immeuble), pour recommander des dépistages sur certains groupes d'habitants. C'est donc une logique opérationnelle. Ils ont des niveaux de gravité qualitativement définis. Le changement de catégorie dans ces évaluations peut déclencher des décisions de libérer des lits en réanimation.

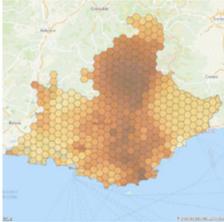
➤ *C'est un outil localisé avec un objectif à très court terme. Il faut aller vite pour prendre des décisions de territoire.*



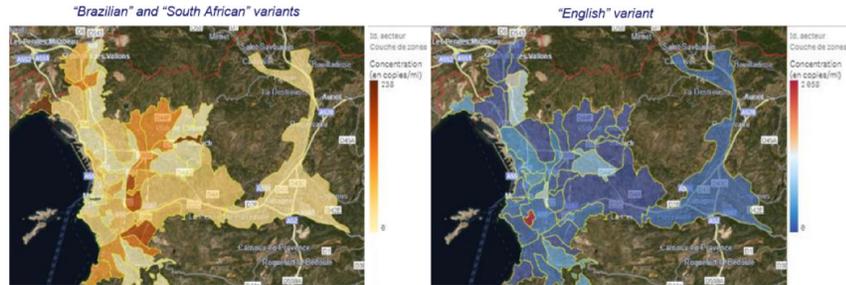
- Multiple explorable online dashboards & cartographies
- Results available in day+1
- Unlimited access 24/7, multi-users
- Documentation available to explain each dashboards & each indicator calculation method (data literacy)
- Intuitive end-user interface



Integration & cross-reference of additional data sources (e.g. population data)



Daily tracking of variants strains in real-time by district (in Marseille, France)



Ils voient aussi l'évolution des proportions des variants. Ils ont ainsi une cinétique rapide sur ce sujet et c'est utilisé à Dunkerque et Nice, par exemple. On se pose la question de surveiller crèches, écoles, maternités de cette façon-là.

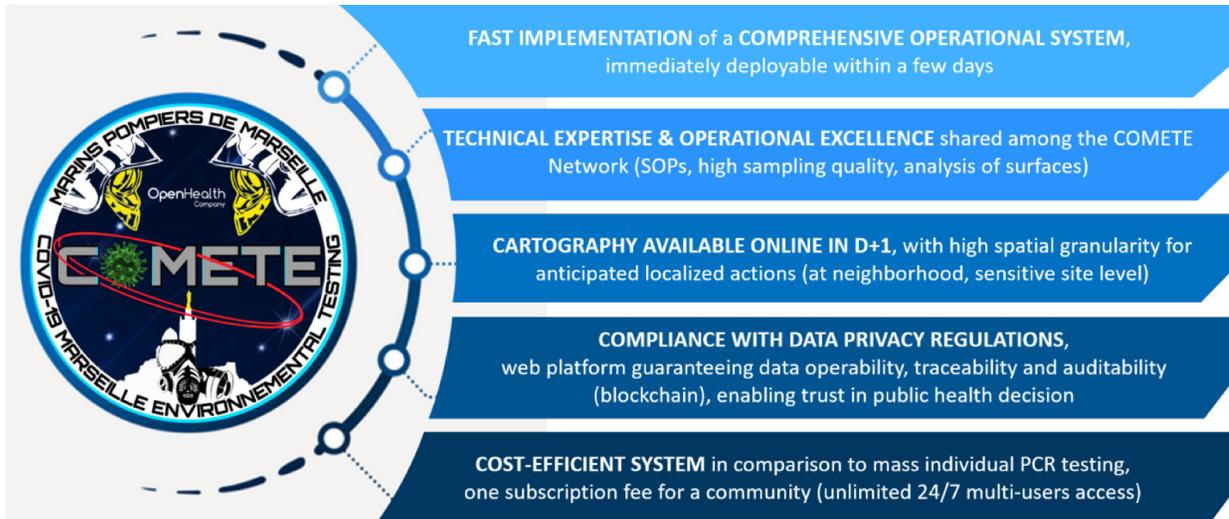
**La sensibilité est forte : on peut identifier un infecté sur 1 000 habitants via les eaux usées. Ces mesures sont donc sensibles.**

Un grand nombre de métropoles du sud de la France est utilisateur de la technique, plus la Bretagne et la Moselle. La Dordogne a rejoint le réseau et s'en sert pour surveiller ses Ehpad et tout son territoire à partir des stations d'épuration.

Le service est déployé sous la forme d'un service web qui permet d'introduire les résultats depuis n'importe où. Sur les gros points de mesure (collecteurs d'eaux usées), ça se fait sous forme de flux automatiques.

## QUELQUES ÉLÉMENTS TECHNIQUES

### CE QU'OFFRE LE RÉSEAU COMETE



### POINTS TECHNIQUES

Plate-forme de data analytics Qlik Sense<sup>101</sup>. Cette plate-forme est interopérable (API, Web services...) et les données sont exportables (portabilité native) à tout moment par l'utilisateur. Ceci est important tant pour la saisie, pour l'interconnexion avec des systèmes d'information des laboratoires que pour l'analyse en aval dans des systèmes plus évolués, ou plus intégrés. C'est le moyen également de partager les données, d'intégrer d'autres sources.

## QUESTIONS D'ORGANISATION

*AB : D'où vient l'initiative ?*

C'est une initiative commune OpenHealth avec les pompiers de Marseille qui ont cette cellule de risques NRBC. Ils ont exploré cette possibilité. Ils ont cherché à voir s'ils pouvaient quantifier l'interprétation de ces mesures. Ils ont trouvé une méthode simple et donc portable qui permet d'avoir le signal opérationnel dont nous avons parlé. Nous avons constaté que notre savoir-faire méritait d'être partagé.

*LG : parvient-on à calibrer le modèle qui fait le lien entre eaux usées et hôpitaux ?*

À Briançon ils ont deux stations d'épuration dont l'une est exposée à la dilution du virus et l'autre pas. On peut s'en servir pour calibrer (à pluviométrie constante).

*AB : avez-vous été aidés ou non ?*

La relation au niveau de la mairie de Marseille a été d'une grande fluidité. La participation au réseau ne demande pas d'adaptation difficile sur le plan technique, mais **il impose une**

101 [https://www.qlik.com/fr-fr/lp/ppc/qlik-sense-business/brand?CampaignID=7013z000000it5d&ppc\\_id=gF59isfl&kw=qlik%20sense&utm\\_content=sgF59isfl\\_pcrd\\_409661229970\\_pmt\\_e\\_pkw\\_qlik%20sense\\_pdv\\_c\\_mslid\\_pgrid\\_12625049054\\_ptaid\\_kwd-68438765356&utm\\_source=google&utm\\_medium=cpc&utm\\_campaign=Qlik\\_France\\_Google\\_Brand\\_DA\\_FR&utm\\_term=qlik%20sense&gclid=Cj0KCQIA-v6yCBhCLARIsABqJTjY7eN4GoaRjzYHI20dROLBxQvOeYIrl31pgv3k7hh5vQIW3ICGvjirMaAt\\_qEALw\\_wcB](https://www.qlik.com/fr-fr/lp/ppc/qlik-sense-business/brand?CampaignID=7013z000000it5d&ppc_id=gF59isfl&kw=qlik%20sense&utm_content=sgF59isfl_pcrd_409661229970_pmt_e_pkw_qlik%20sense_pdv_c_mslid_pgrid_12625049054_ptaid_kwd-68438765356&utm_source=google&utm_medium=cpc&utm_campaign=Qlik_France_Google_Brand_DA_FR&utm_term=qlik%20sense&gclid=Cj0KCQIA-v6yCBhCLARIsABqJTjY7eN4GoaRjzYHI20dROLBxQvOeYIrl31pgv3k7hh5vQIW3ICGvjirMaAt_qEALw_wcB)

**obligation de partage des résultats.** On procède à des essais pour qualifier un labo et son personnel sur la méthode. Une fois qu'ils ont maîtrisé la méthode, ils peuvent l'adapter aux machines déployées sur leurs grands sites d'épuration et collecteurs. Toujours avec une obligation de partage.

Nous n'avons pas eu besoin de gros moyens spécifiques, donc pas besoin de mobiliser des financements.

*LG : liens entre analyses pour animaux versus pour humains ?*

En médecine vétérinaire, les raisonnements épidémiques sont permanents. Les pandémies sont là en permanence. Il y a donc déjà la culture, pour les tests et les mesures en particuliers. Ainsi PCR y est très connu depuis longtemps.

**Les capacités d'analyse en médecine vétérinaire sont très largement supérieures à celles de la médecine humaine (facteur 100), mais cette capacité n'a pas été sollicitée lors de la pandémie covid. Les mêmes tests qui ont coûté très cher pour le covid existaient déjà pour rien chez les vétérinaires.**

*LG vos relations avec Obépine ?*

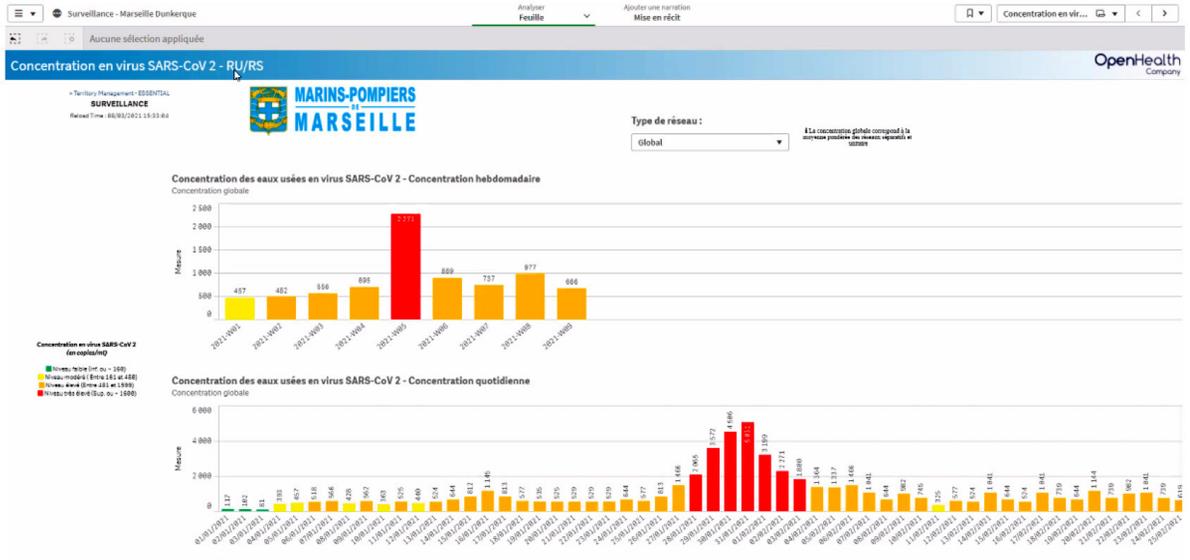
Le contact a été tenté mais pas établi. *Obépine* semble plus d'inspiration recherche. Mais les deux doivent être combinables.

Le cas des tests vétérinaires et des agréments pour tel ou tel labo : les établissements vétérinaires ont proposé leur contribution aux tests, mais n'ont pas été agréés. Nous avons tenté cet agrément en avril 2020.

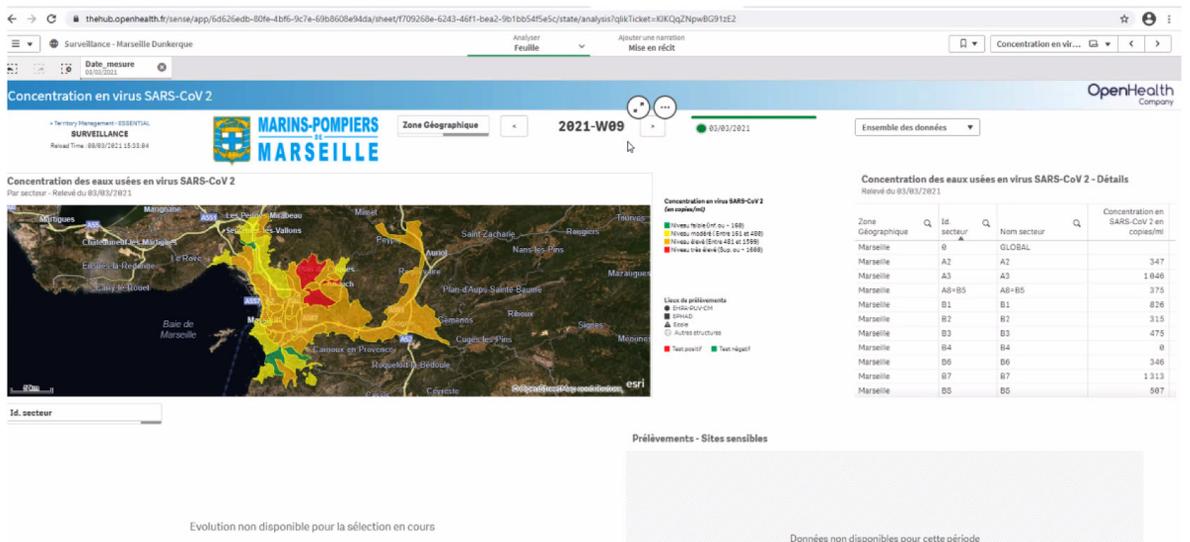
Parfois, l'augmentation des résultats positifs a été expliquée par l'augmentation des tests, en totalité. Le changement est expliqué par le changement de thermomètre.

## DÉMONSTRATION

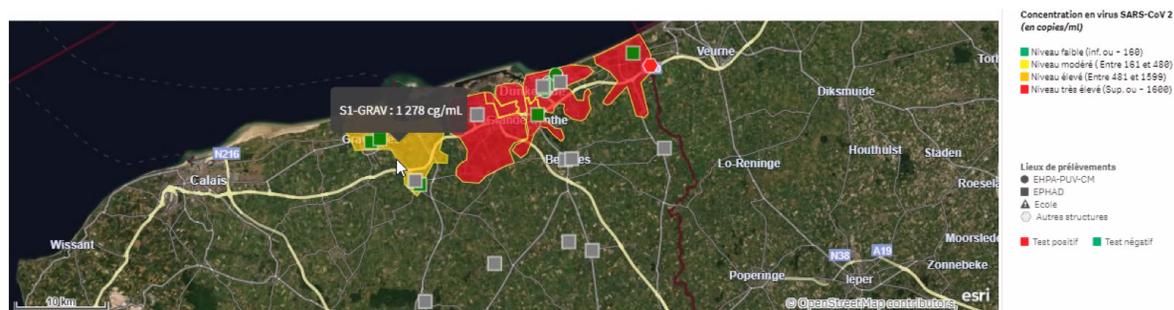
### DONNÉES MARSEILLE, CAPTURES D'ÉCRAN. INDICATEUR BAR CHART.



On détecte les phases ascendantes pour informer les services d'urgence de ne pas accepter la charge autre que covid. On mesure des traces de virus, pas des gens malades. Mais les deux sont liés statistiquement. L'anticipation par rapport à une surcharge du système de santé est de 1 semaine environ (c'est valable partout). C'est donc utilisable dans un territoire.



Ils peuvent ainsi concentrer les moyens dans les secteurs rouges (ici Marseille nord). Usage très opérationnel. Ils peuvent remonter jusqu'aux immeubles, voire étages ils préviennent les gens.



Sur la même carte on juxtapose stations d'épuration et entités sensibles (Ehpad...). On peut ainsi prévenir tel ou tel Ehpad. On voit un Ehpad vert en zone rouge. On focalise sur un point positif ça ouvre un lien vers le détail du test qui explique ce point vert.

## QUELQUES LEÇONS

### CONCERNANT L'ORGANISATION

**C**oncernant l'harmonisation et l'articulation de ces actions locales avec les règles nationales, une bonne coordination locale et beaucoup d'agilité sont requises. Les territoires qui mettent cela en place sont ceux qui sont pragmatiques et savent gérer ce genre de souplesse. Les ARS en revanche, travaillent de manière très solitaire et inféodée à Paris. C'est différent des pompiers qui sont habitués à l'agilité. Plus d'agilité serait bienvenu et utile.

### CONCERNANT LA TRAÇABILITÉ

**B**eaucoup de données ont circulé, mais avec des indicateurs qui ont de gros biais. Plus on agrège les indicateurs, moins ils ont un sens lié aux données primitives, et plus les biais sont pervers.

Il faut, en fait, la traçabilité totale du kit et de la machine utilisés la traçabilité totale est clé. Dans le covid-19 la méthode de comptage a changé trois fois. Ces biais ont été renforcés par le système d'information. L'entrée des informations était : + ou -. Et donc d'où venait le + ou le - était perdu. Or, la décision pouvait dépendre du kit ou de la méthode d'interprétation. C'est le problème avec les synthèses en arborescence où l'on abstrait de plus en plus.

## **ANNEXE 7 – ENTRETIEN AVEC PATRICK LAGADEC**

*(2021/01/08)*

*Notes de Laurent Gouzènes, Revisés par Patrick Lagadec*

## INTRODUCTION

**A**ncien chercheur à l'École polytechnique, docteur d'État en sciences politiques, Patrick Lagadec est spécialiste mondial de la gestion des crises et consultant. (Contact : [Patrick@patricklagadec.net](mailto:Patrick@patricklagadec.net))

À ce titre il dispose d'une expérience considérable et mondiale, ainsi que la disponibilité d'un savoir théorisé/modélisé sur la gestion de crises.

Il a réalisé de nombreux ouvrages et analyses sur les crises passées et documents (cf bibliographie), dont nous conseillons fortement la lecture. Il propose une méthodologie fondée sur la trilogie : se préparer, réagir avec les bons outils, capitaliser l'expérience pour la prochaine crise.

Le point le plus essentiel dans son message est que **nous sommes désormais projetés dans des univers de risques et de crises largement étrangers aux univers des années 1980-2000 quand ont été forgés nos cartographies des risques et nos logiques de gestion des crises. Et cette sortie de nos « domaines de vol », exige désormais une nouvelle approche particulièrement inventive sur nombre de dimensions – à commencer par l'exercice du leadership, les démarches d'aide à la décision, les dynamiques de coopération au sein des systèmes organisationnels et sociaux.**

Les questions abordées par Patrick Lagadec sont entre autres :

- « Comment piloter les grandes situations de crise et de rupture qui, en raison des turbulences et dislocations du monde actuel, vont se multiplier et muter dans leur envergure comme dans leur nature ?
- Comment préparer les dirigeants pour qu'ils puissent assumer leurs responsabilités quand ils se retrouvent confrontés à cette « matière noire déboussolante » de la crise majeure dans nos univers désormais chaotiques et inconnus ? »

## L'ATTITUDE FACE À LA CRISE

**O**n est en situation gravissime où n'importe quelle décision qu'il faudrait prendre est inaccessible. Pourtant il y a plein de conférences où les méthodes et réponses sont exposées.

La crise par essence paralyse la réflexion (« personne n'aurait imaginé un truc pareil »). On va percevoir très vite que l'on sera sollicité en dehors de sa zone de confort. Dans la pratique, réfléchir à l'avance sur les failles est éprouvant.

*La psychologie des décideurs est donc un point clé de la gestion de la crise.*

La difficulté est la confrontation à des sujets pour lesquels on ne dispose pas de réponse. **Il faudrait donc outiller et préparer pour réagir dans l'inconfort, et en univers inconnu.**

Le problème n'est pas de trouver les meilleures cordes pour faire le *Mont Blanc* : il faut ne pas monter en espadrilles ! Le « on » désigne tout le système : constat de vulnérabilité, refus de préparer la logistique à l'avance, de demander le consentement à l'avance. D'où la viscosité lourde de tous les systèmes de décision.

## QUELLES ORGANISATIONS À METTRE EN ŒUVRE ?

J'essaie autant qu'il est possible de donner des pistes. Voir *Le continent des imprévus* qui donne l'essentiel. J'ai aussi écrit Sécurité nationale. Navigabilité des organisations en univers chaotique. C'est sur mon site et tout est accessible et gratuit. J'y ai fait une analyse covid au début et tracé des pistes de réflexion-action.

L'ornière habituelle : une structuration en silo qui se rigidifie par réflexe au premier signe d'une sortie des univers convenus. Le défi : aussi longtemps que chaque silo est plus enclin à se barricader qu'à ouvrir des voies inventives avec d'autres, la crise se rend maîtresse du terrain.

*Du point de vue pratique, les technologies de l'information, sont-elles un élément de réponse ou un accélérateur de crise ?*

Il faut savoir se configurer rapidement sur une nouvelle crise.

## DES ATTITUDES SPÉCIFIQUES À LA FRANCE ?

C'est un problème plus culturel. Le sujet fait tellement peur qu'il est inacceptable : proposer une réflexion en dehors des cadres est le plus souvent rejeté, avec mise à l'écart des personnes qui osent pointer des questions pour lesquelles les réponses convenues ne suffiront pas. Le sujet est doublement difficile : intellectuellement insupportable, psychiquement intolérable. D'où l'ouverture des parapluies, les mises en mode « tortue » de défense – qui se verront rapidement englouties. Le principe même de la réflexion est ainsi régulièrement refusé.

*LG : quid des conventions de citoyens ? Pourquoi pas sur la covid-19 ?*

Oui, il va falloir intégrer de multiples acteurs dans les processus et l'inventivité est de rigueur. Mais j'insisterai, pour que l'ouverture soit possible (et pas sur un mode panique), sur la nécessité d'avoir institutionnalisé des capacités de prise de recul, d'intelligence stratégique temps réel. C'est l'idée de la démarche de **Force de Réflexion Rapide (FRR)**.

### QUELQUES EXPÉRIENCES :

- Sur l'ouragan Katrina, voir l'expertise hors pair de l'amiral Thad Allen<sup>102</sup> pour un retour d'expérience très intéressant.
- Nous avons animé une journée entière sur *un Tsunami à Nice*. On a discuté les hypothèses qui permettraient de maintenir les services essentiels (pas de coupure électrique, l'aéroport marche...). La réponse a été : ne posez pas d'hypothèses qui ne correspondent pas aux conditions pour lesquelles nous avons les solutions. On l'a vu aussi pour la tempête IRMA. La viscosité est exacerbée par la confrontation à l'ingérable.
- Dans le passé chez EDF, sous l'impulsion de Pierre Bérout alors en charge de la maîtrise des risques, et avec Xavier Guilhou, nous avons développé cette FRR mentionnée précédemment : un petit groupe de gens d'une trentaine de personnes habituées à chercher les signaux faibles et plus encore aberrants. En 2020 par exemple, les réticences initiales de

102 <https://hal.archives-ouvertes.fr/hal-00243057>

- l'OMS auraient dû alerter, loin de « rassurer ».
- Pareille méthode, institutionnalisée, avec équipes formées, aurait été extraordinairement utile sur tout le processus de conduite de la covid-19. Ne pas avoir cet outil qui oblige à s'interroger en anticipation, en temps réel comme sur le temps long, **sur les quatre questions clés d'une FRR : le « de quoi s'agit-il vraiment ? », les « pièges », les « acteurs », les « combinaisons d'impulsions inventives »**, condamne à le plus souvent arriver trop tard, à tomber dans les ornières les plus pénalisantes, à perdre crédibilité et légitimité.
  - J'ai vu aussi le cas lors de la crise 2008 sur les *subprimes* : c'était un sujet tabou depuis quelque temps.

## BIBLIOGRAPHIE

On pourra lire très utilement les références suivantes :

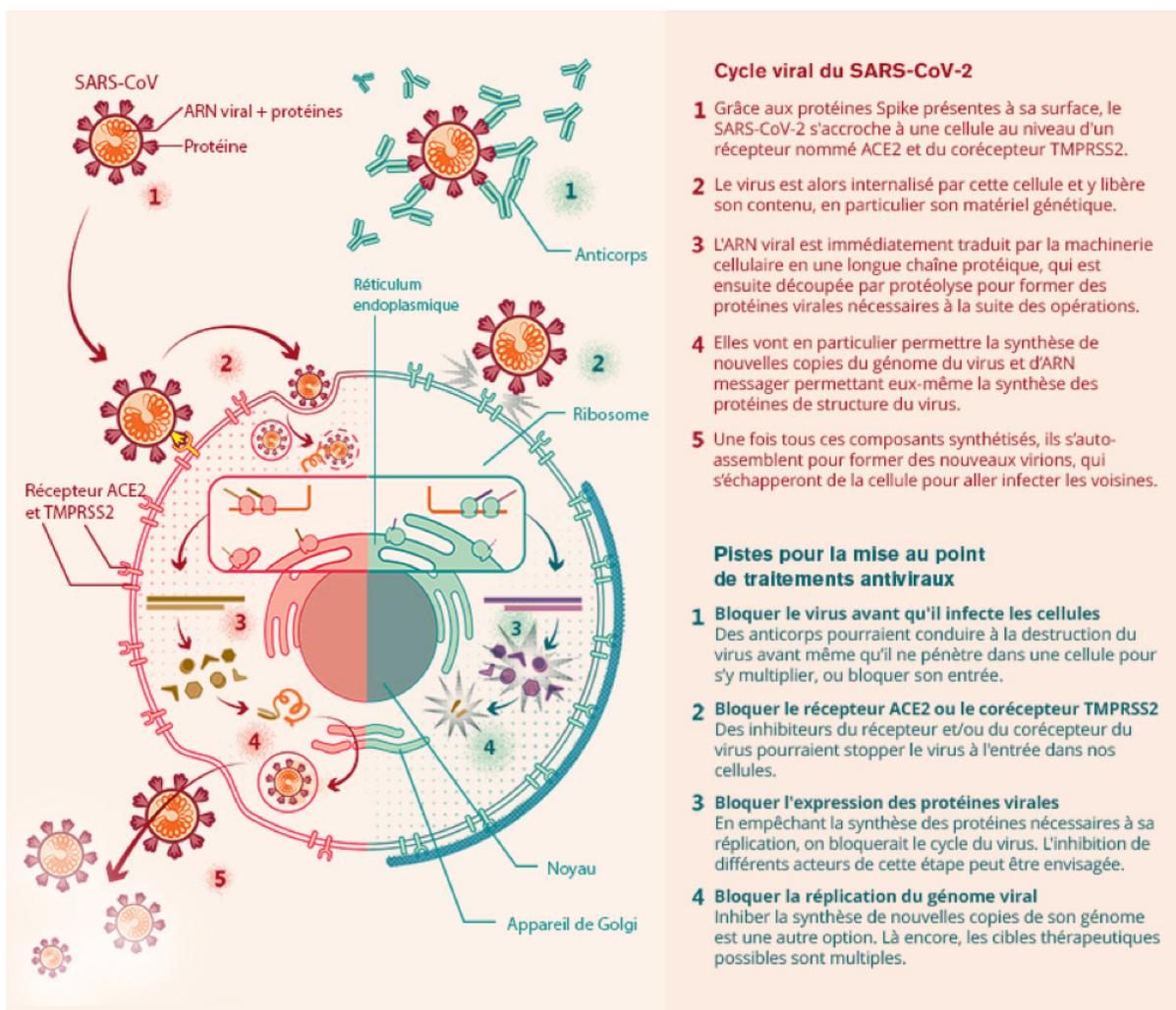
- Site de P Lagadec : <http://www.patricklagadec.net/fr>
  - Crises « hors cadres » : oser un enseignement
  - [Ne pas rater le retour d'expérience](#)
- **Livres :**
  - [Le continent des imprévus](#), Mantoba, Les Belles lettres 2015
  - [Le temps de l'invention](#) (2019)
- **Articles :** Une série d'articles dans la revue « [Crisis Response](#) », sur LinkedIn (<https://www.linkedin.com/in/patrick-lagadec-b2305714/detail/recent-activity/posts/> )

## ANNEXE TECHNIQUE ET BIBLIOGRAPHIQUE :

# 1. État de la pandémie

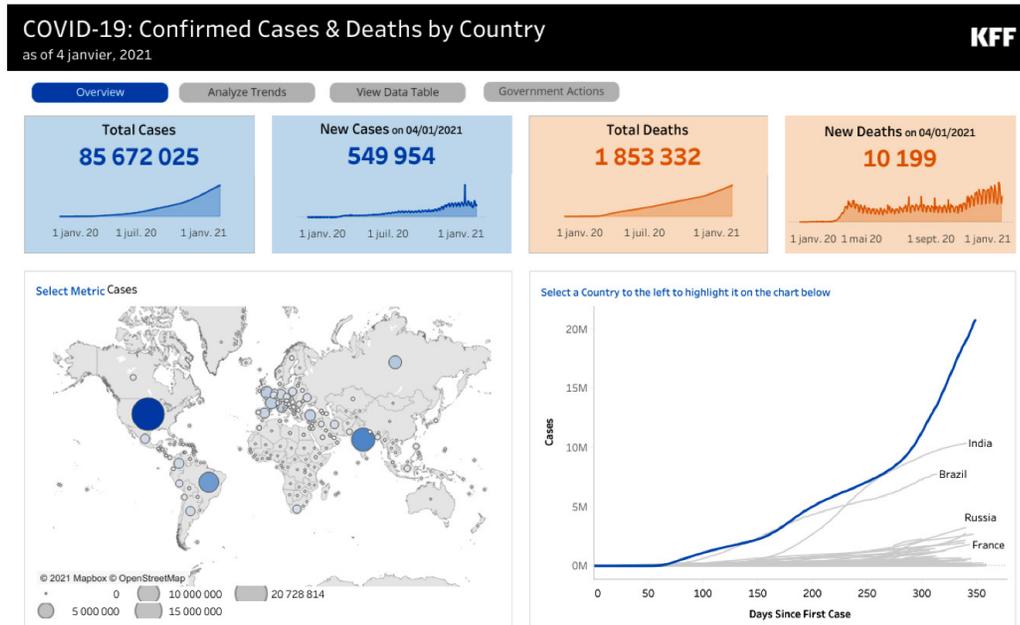
## PRÉAMBULE

<https://www.inserm.fr/information-en-sante/dossiers-information/coronavirus-sars-cov-et-mers-cov> Inserm, schéma introductif sur le SARSCOV-2. On trouve sur ce lien une présentation introductive dont est extraite la figure ci-dessus.



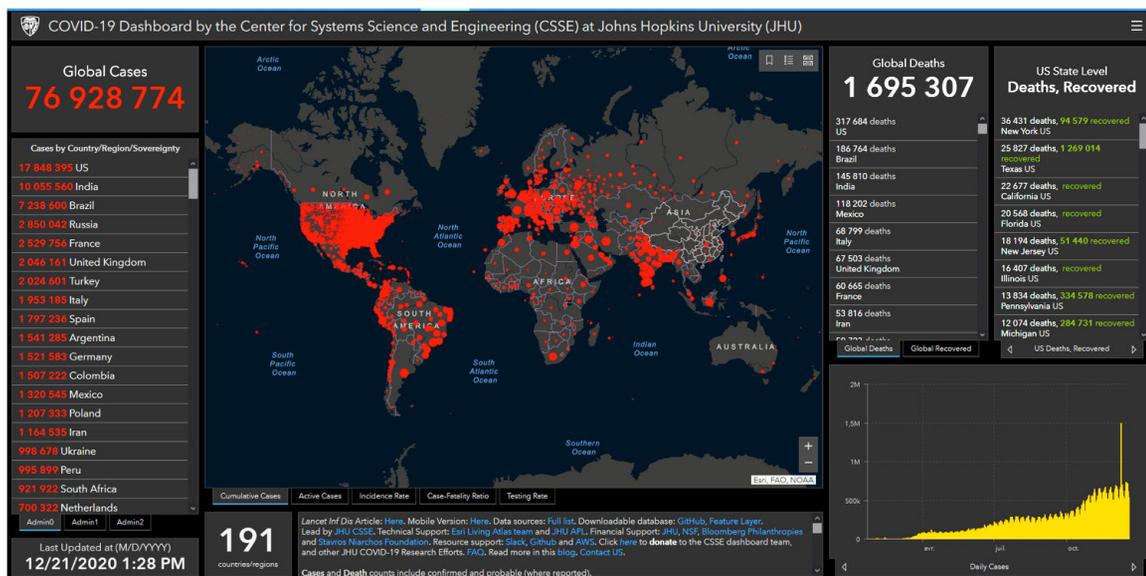
## LIENS BIBLIOGRAPHIQUES SUR LE SUIVI

- Site <https://www.kff.org/coronavirus-covid-19/fact-sheet/coronavirus-tracker/>



Source: Johns Hopkins University (JHU) Coronavirus Resource Center; last updated with data from 04/01/2021. Data prior to January 22, 2020 are from the World Health Organization's (WHO) Coronavirus disease (COVID-2019) situation reports. Notes: Cumulative case totals include both laboratory confirmed and clinically diagnosed cases; prior to February 14, 2020, totals include only laboratory confirmed cases. Japan's totals include cases that have been identified on the Diamond Princess cruise ship (except in cases that have been re-categorized by a reporting country).

Dans ce diagramme, la courbe en gras correspond à la France.



- <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/covid-19-etat-des-connaissances-et-veille-documentaire>  
Site de Santé Publique France. Le document suivant contient une liste de publications avec résumés et liens :: [https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/covid-19-etat-des-connaissances-et-veille-documentaire/articles/epidemie-de-coronavirus-le-role-de-la-veille-scientifique-et-documentaire/blocs/2020-12-17\\_covid-19\\_weekly\\_document\\_list](https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/covid-19-etat-des-connaissances-et-veille-documentaire/articles/epidemie-de-coronavirus-le-role-de-la-veille-scientifique-et-documentaire/blocs/2020-12-17_covid-19_weekly_document_list)
  - [https://opendata.idf.inserm.fr/cepfdc/covid-19/certification\\_electronique.html](https://opendata.idf.inserm.fr/cepfdc/covid-19/certification_electronique.html)  
Ce lien indique les décès COVID certifiés électroniquement par l'Inserm
  - [https://en.wikipedia.org/wiki/Template:COVID-19\\_pandemic\\_data#covid-19-pandemic-data](https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data#covid-19-pandemic-data)  
Wikipedia : tableau mondial de la pandémie maintenu journallement.
  - <https://medium.com/wintoncentre/why-i-am-in-a-higher-priority-group-for-a-vaccine-than-younger-people-with-chronic-health-974621eec30>  
Article dont nu co-auteur est David Spiegelhater, statisticien renommé très reconnu dans la communauté IA & statistiques
  - <https://presse.inserm.fr/covid-19-un-modele-mathematique-pour-ameliorer-le-systeme-de-depistage-francais-et-mieux-controler-lepidemie/41798/?amp=1>  
Inserm voir aussi l'article dans la revue Nature pointé dans l'article ci-dessus
- « Une forte sous-estimation des cas est un frein majeur au contrôle de l'épidémie. En effet, les cas non identifiés continuent de propager le virus mettant à risque notre stratégie de contrôle. Tester-tracer-isoler est le seul moyen dont nous disposons pour continuer à freiner la propagation du virus en allégeant les mesures restrictives ciblant toute la population. La tâche est d'autant plus ardue que nous sommes en période hivernale. Notre étude suggère qu'il faut continuer à renforcer les capacités de dépistage pour qu'elles soient plus ciblées, efficaces et accessibles à tous pour lutter contre la pandémie »,*
- conclut Vittoria Colizza.*
- <https://fr.statista.com/statistiques/1091585/morts-infections-coronavirus-monde/> Nombre d'infections dues au coronavirus par pays du monde 16 décembre 2020  
Publié par [E. Moyou](#), 18 déc. 2020. Cette statistique montre le nombre de personnes infectées par le [coronavirus COVID-19](#) dans le monde au <https://www.worldometers.info/coronavirus/> 2020, selon le pays. On observe ainsi que sur 73,8 millions d'infections à ce jour, 86.770 étaient localisées en [Chine](#), qui était initialement le foyer mondial de l'épidémie. Or, depuis la fin du mois de mars, ce sont les États-Unis qui comptabilisent le plus grand nombre de personnes contaminées et de [décès](#) dus au COVID-19.

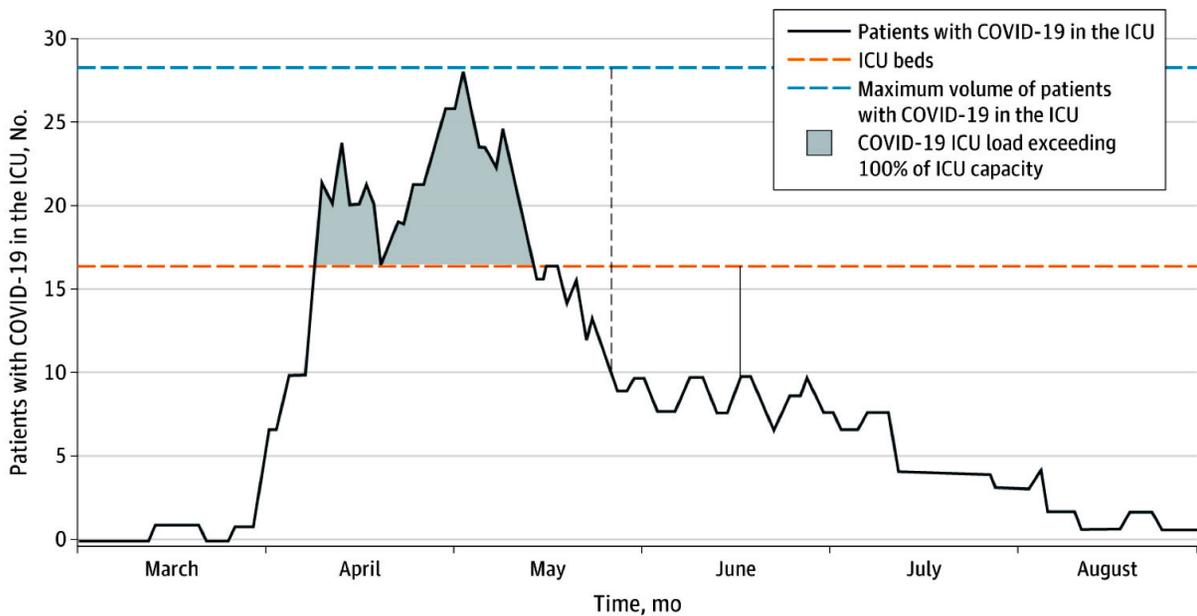
- <https://www.china-briefing.com/news/china-coronavirus-updates-latest-developments-business-advisory-part-2/>

*Semble être une agence officielle collecte les données journalières dans un format texte.*

- Worldometer (données mondiales collectées journallement)  
<https://www.worldometers.info/coronavirus/worldwide-graphs/>  
<https://www.worldometers.info/coronavirus/#countries>  
<https://www.worldometers.info/coronavirus/country/china/>

- [https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2775236?utm\\_source=silverchair&utm\\_campaign=jama\\_network&utm\\_content=covid\\_weekly\\_highlights&utm\\_medium=email](https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2775236?utm_source=silverchair&utm_campaign=jama_network&utm_content=covid_weekly_highlights&utm_medium=email)

*Jama Network Open journal (Nov 2020): effet de la surcharge hospitalière*



Although strain on hospital capacity has been associated with increased mortality in nonpandemic settings, studies are needed to examine the association between coronavirus disease 2019 (COVID-19) critical care capacity and mortality. The results suggest that the risk of mortality would be highest if a COVID-19 patient’s stay was during the peak of ICU demand and if ICU caseload approached or exceeded ICU bed capacity.

## VACCINS

- <https://berthub.eu/articles/posts/reverse-engineering-source-code-of-the-biontech-pfizer-vaccine/>
- <https://renaudguerin.net/posts/partie-2-explorons-le-code-source-du-vaccin-biontech-pfizer/> pour la version en français.  
*Article explicatif sur le vaccin BioNTech/Pfizer, plus particulièrement adressé à des lecteurs de culture informaticienne tout-à-fait passionnant pour un lecteur de formation scientifique quelle qu'elle soit.*
- [https://theconversation.com/avantages-desavantages-risques-ce-quil-faut-savoir-sur-les-vaccins-a-arn-152333?utm\\_medium=amplinkedin&utm\\_source=linkedin](https://theconversation.com/avantages-desavantages-risques-ce-quil-faut-savoir-sur-les-vaccins-a-arn-152333?utm_medium=amplinkedin&utm_source=linkedin)  
*Complète le lien précédent par un article moins technique, mais intéressant pour ce qui concerne les avantages et risques liés à la nouvelle technologie de vaccins à ARM messenger.*

## MORTALITÉ ET SURMORTALITÉ

Les deux affirmations suivantes sont justes simultanément :

- le covid a fait 70 000 morts ;
- la surmortalité du covid est de 35 000 morts.

### **EXPLICATION :**

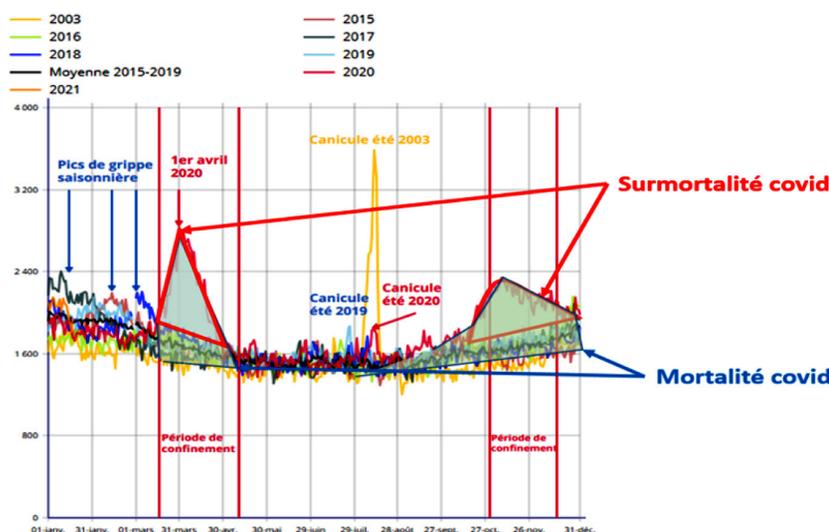
Le décompte des 70°000 morts est un **Comptage individuel**, qui enregistre la cause première du décès. Par exemple, un malade du cancer, même sans issue sur le court terme, qui est atteint par la covid-19 et qui en meurt suite à des problèmes respiratoires, meurt de la covid, pas du cancer (le cancer devient une comorbidité). Du coup, le nombre de décès par cancer va aussi baisser, sauf si on compte ces comorbidités correctement. On ne peut pas pour autant dire que 70 000 personnes auraient eu la vie sauve si la covid n'avait pas existé.

La surmortalité est un **calcul statistique**, qui prend en compte les années passées et produit une différence (là il ne faut pas oublier non-plus l'évolution de la moyenne). Ainsi le malade du cancer qui décède plus tôt que prévu ne change pas la mortalité, calculée statistiquement. Par contre un individu qui décède brusquement de la covid est compté dans la surmortalité. Le nombre de décès quotidien prévisible résulte de deux composantes :

- une composante liée à la démographie : les classes d'âge issues du baby-boom de 1945 atteignent en 2020 les 75 ans : la mortalité globale augmente au rythme de cette nouvelle structure de population (soit entre + 600 à + 9000 par an). Il est en particulier très incorrect de parler de la « moyenne de 2015 à 2019 » sans prendre en compte cette tendance haussière ;

- une composante de fluctuation saisonnière (le rythme quotidien des décès en hiver est en moyenne > 15 % de celui de l'été).

Une estimation graphique, qui se raccroche aussi aux données statistiques donne une surmortalité d'environ 35 000 personnes en 2020 :

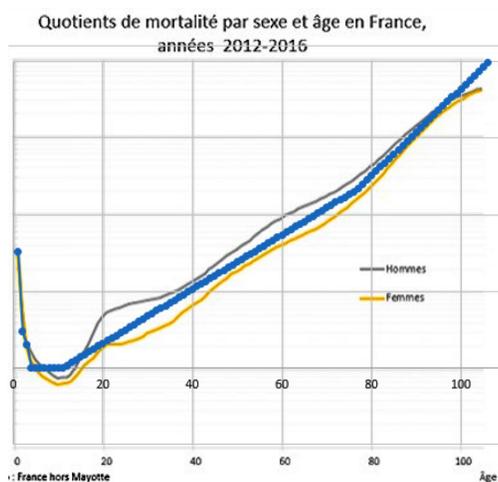


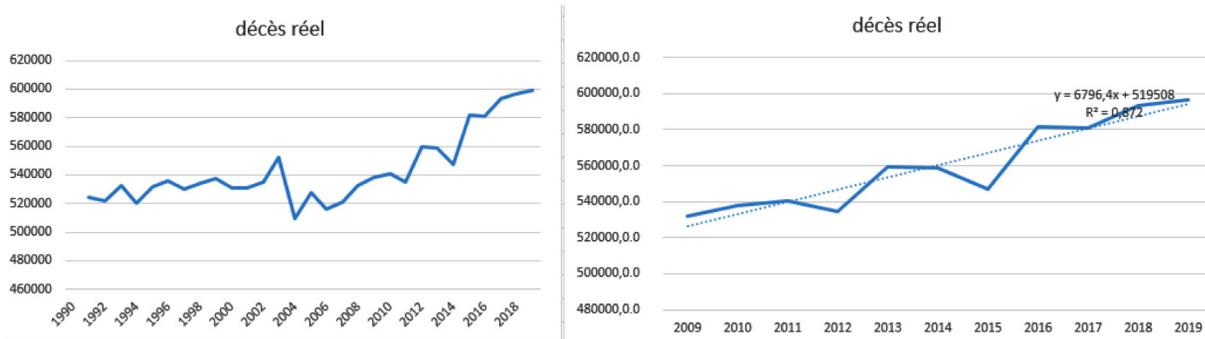
Les mesures de santé publique exceptionnelles se justifient lorsqu'il y a surmortalité, c'est-à-dire lorsque les courbes de mortalité courante sortent de l'intervalle de variations statistiques courantes. Évidemment, le point dur de la gestion de cette épidémie (au début) est que le dispositif d'alerte standard devait prendre en compte l'évolution rapide de l'épidémie (rappel : x 10 tous les 10 jours au début de mars 2020), et donc anticiper cette anomalie. En valeur absolue, la mortalité en 2020 (provisoire) augmente à 667.400 décès toutes causes confondues, 53.900 de plus qu'en 2019 (+9 %), selon un premier bilan établi par l'Insee. La pyramide des âges en France traduit le baby-boom de 1945 : les générations nombreuses post guerre rentrent maintenant dans leur 75ème année.

Or le taux de mortalité varie considérablement en fonction de l'âge, figure ci-contre.

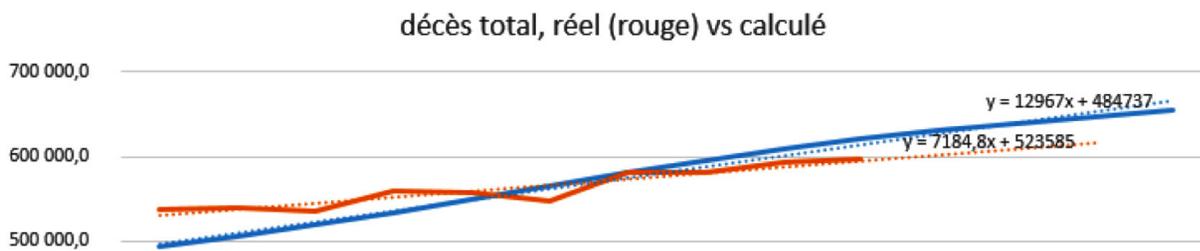
Ceci nous amène à prédire une augmentation de la mortalité 'naturelle' dans les prochaines années, liée à l'augmentation considérable du nombre d'anciens.

La courbe historique des décès montre ainsi une augmentation récente mais à tendance régulière d'augmentation des décès (source Insee), sur une pente observée de + 6800 décès par an :





(Note : la courbe « théorique » calculé à partir de la pyramide des âges et de la table de mortalité donne ~+13000 décès par an.) :



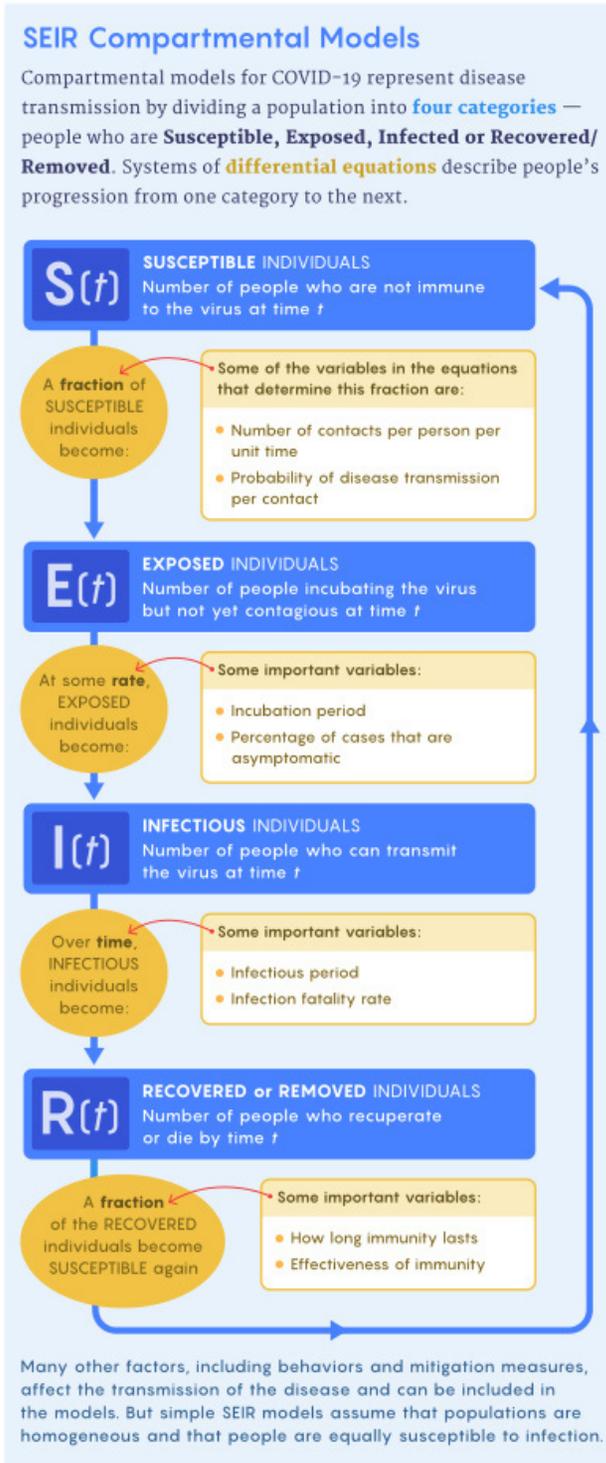
On constate qu'un retard d'environ 5000 décès annuels a été pris depuis 3 ans entre ces deux courbes, soit un déficit d'environ 15000 décès auxquelles il faut ajouter la dernière année, soit 5000 décès. Il semble qu'une part de la population âgée, qui avait été épargnée par des épisodes de grippe des dernières années, ait été ravagée par le covid-19. La surmortalité serait alors d'environ  $53900 - 20000 = 33900$ , cohérente avec l'analyse purement graphique présentée en premier lieu.

## ANNEXE TECHNIQUE ET BIBLIOGRAPHIQUE :

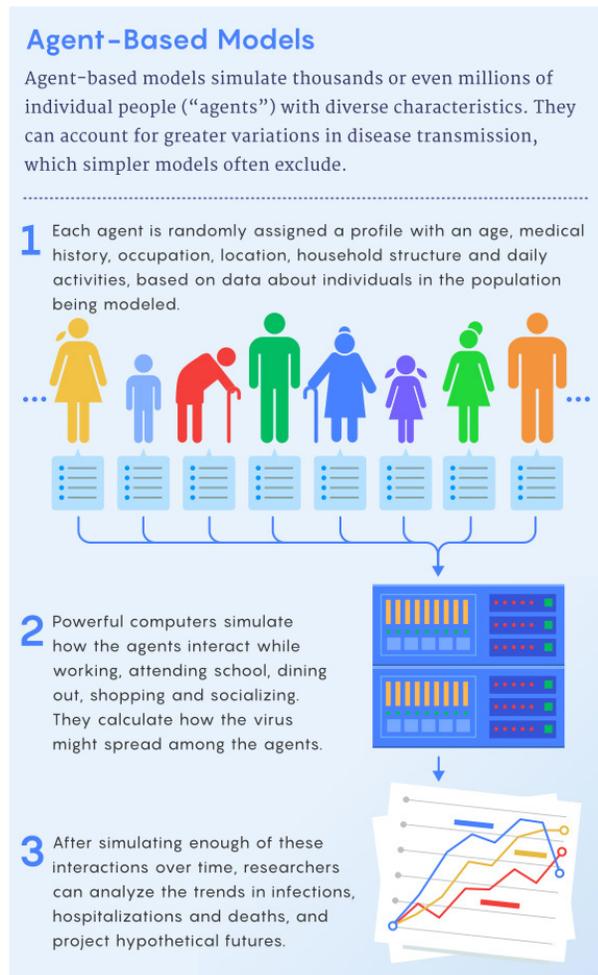
# 2. Sur la modélisation épidémiologique

## SUR LES GRANDES FAMILLES DE MODÉLISATION

- Source <https://www.quantamagazine.org/the-hard-lessons-of-modeling-the-coronavirus-pandemic-20210128/#>



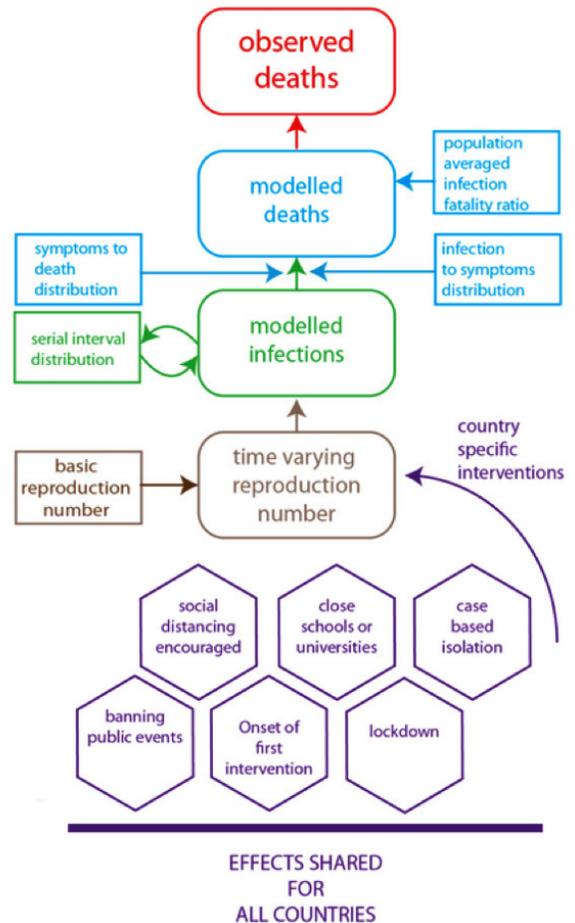
Ces figures indiquent par des schémas les deux grandes familles de modélisation.



## SUR L'AJUSTEMENT DES MODÈLES SUR LES DONNÉES

**Source:** [Flaxman, Mishra, Gandy et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe." Nature, may 2020. https://www.nature.com/articles/s41586-020-2405-7](https://www.nature.com/articles/s41586-020-2405-7)

- We use a Bayesian model that calculates backwards from the deaths observed over time to estimate transmission that occurred several weeks prior.
- Changes in the reproductive number – a measure of transmission - are an immediate response to interventions.
- Each intervention has the same effect on the reproduction number across countries and over time, excluding the lockdown intervention. For lockdown, we now estimate a global effect and a country specific effect.
- The interventions that we consider have different implementation details across countries, which we do not take into account.
- We make various further assumptions that our model results are contingent on (e.g. about the time between infections, about being able to observe all COVID-related deaths, about the time between infection and death and about the infection-fatality-ratio).



Data source: Daily time series data from ECDC. Interventions related data can be found [here](#).

More details can be found in the [original publication](#), in the [technical model description](#) and in the [source code](#) of the model.

Extrait de la page Web qui accompagne cet article, qui présente leur outil d'ajustement de leurs modèles épidémiologiques par apprentissage Bayésien sur les données collectées, avec application à l'Europe. La reproduction d'une page de la publication correspondante (ci-après) donne des indications sur la méthode suivie pour la collecte de données ainsi que l'ajustement des modèles. La figure qui suit avec les courbes de mortalité provient du même article.

## Article

### Methods

#### Data

Our model uses daily consolidated death data from the ECDC for 11 European countries currently experiencing the COVID-19 epidemic: Austria, Belgium, Denmark, France, Germany, Italy, Norway, Spain, Sweden, Switzerland and the UK. The ECDC provides information on confirmed cases and deaths attributable to COVID-19. For population counts, we use the United Nations Population Division age-stratified counts<sup>18</sup>.

We also catalogue data on the nature and type of major non-pharmaceutical interventions. We looked at the government webpages from each country as well as their official public health webpages to identify the latest advice or laws being issued by the government and public health authorities. We collected the following: school closure ordered; case-based measures; public events banned; social distancing encouraged; lockdown decreed; and the time of the first and last intervention. A full list of the timing of these interventions and the sources we have used is provided in the Supplementary Notes, Supplementary Table 2.

By using the ECDC data, we rely on a consolidated data source compiled by the ECDC, who include many sources of data each day, constantly refining and updating data using a comprehensive and systematic process. However, despite the rigorous protocols, countries may vary in the specifics of the data that they report to the ECDC. For example, there is variation in reporting (that is, community versus hospital) and time lags. Despite these issues, we use ECDC data to ensure as much consistency as possible across all countries.

#### Model

A visual summary of our model is presented in Extended Data Fig. 3; details are provided in the Supplementary Methods.

We fit our model to observed deaths according to ECDC data from 11 European countries. The modelled deaths are informed by an infection-to-death distribution (Supplementary Fig. 1; derived from assumptions about the time from infection to the onset of symptoms and about the time from the onset of symptoms to death), and the population-averaged infection fatality ratio (adjusted for the age structure and contact patterns of each country, as discussed in the Supplementary Methods, Supplementary Table 3).

Given these distributions and ratios, modelled deaths are a function of the number of infections. The number of infections is modelled as the product of  $R_t$  with a discrete convolution of the previous infections. Individual components of this convolution sum are weighted by the generation time distribution (the average time from the infection of one person to the time at which they infect another; Supplementary Fig. 2). In our work, we approximate the generation time distribution using the serial interval distribution.  $R_t$  is a function of the initial  $R_t$  before interventions and the effect sizes from interventions, in which interventions are modelled as piecewise constant functions.

Following the Bayesian hierarchy from bottom to top gives us a full framework to see how interventions affect infections, which can result in deaths. A schematic of our model is shown in Extended Data Fig. 3. To maximize the ability to observe the effect of interventions on deaths, we fit our model jointly for all 11 European countries, and use partial pooling of information between countries with both individual and shared effects on  $R_t$ . Partial pooling operates on the last intervention, which is—in most cases—lockdown. The effect of partial pooling can be seen in Supplementary Discussion 12, Supplementary Fig. 29. We chose a balanced prior that encodes the prior belief that interventions have an equal chance of having an effect or not, and ensure a uniform prior on the joint effect of all interventions (Supplementary Fig. 3). We evaluate the effect of our Bayesian prior distribution choices and evaluate our Bayesian posterior calibration to ensure our results are statistically robust.

We perform extensive model validation and sensitivity analyses. We validate our model by cross-validation over a 14-day period (Supplementary Discussion 1, Supplementary Table 1) and we show the fits for holdout samples in Supplementary Figs. 5–15. We check the convergence of the Markov chain Monte Carlo sampler (Supplementary Fig. 4). We consider the sensitivity of our estimates of  $R_t$  to the mean of the generation distribution (Supplementary Discussion 3, Supplementary Figs. 16, 17). We further show that the choice of generation distribution does not change our counterfactual conclusions (Supplementary Fig. 18). Using univariate analyses and uninformative priors, we find (Supplementary Fig. 19) that all effects on their own serve to decrease  $R_t$  (Supplementary Discussion 4). We compare our model to a non-parametric Gaussian Process model (Supplementary Discussion 5). To assess the effect of individual countries on the results, we perform a ‘leave one country out’ sensitivity analysis (Supplementary Discussion 6, Supplementary Figs. 20, 21). To validate our starting values of  $R_t$ , we compare our model against an exponential-growth linear model (Supplementary Discussion 7, Supplementary Fig. 22). Instead of a joint analysis, we consider fits of our model to individual countries (Supplementary Discussion 8, Supplementary Figs. 23–26). We perform a sensitivity analysis with respect to the onset-to-death distribution (Supplementary Discussion 9, Supplementary Fig. 27). We validate our probabilistic seeding scheme through an importance-sampling leave-one-out cross-validation (Supplementary Discussion 10). We consider a model extension with a constant, probabilistic under-reporting (Supplementary Discussion 11), finding that  $R_t$  does not change substantially (Supplementary Fig. 28).

Our model is different to other approaches (such as EpiEstim<sup>19</sup>) that use the discrete renewal equation. We use the renewal equation as a latent process to model infections and propose a generative mechanism to connect these infections to death data. Simply applying the renewal equation directly to death data requires positing a mechanism in which deaths in the past can cause future deaths (see, for example, ref.<sup>20</sup>). In addition, for  $R_t$ , we are able to use a functional relationship in which non-pharmaceutical interventions can have a direct effect on  $R_t$ .

#### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

#### Data availability

Death counts for the 11 European countries for the time period in our study and the full set of posterior draws from our model are available at <https://reshare.ukdataservice.ac.uk/854380/>.

#### Code availability

All source code and data necessary for the replication of our results and figures are available at <https://github.com/ImperialCollegeLondon/covid19model>. An R package based on our method is available at <https://imperialcollegelondon.github.io/epidemta/>.

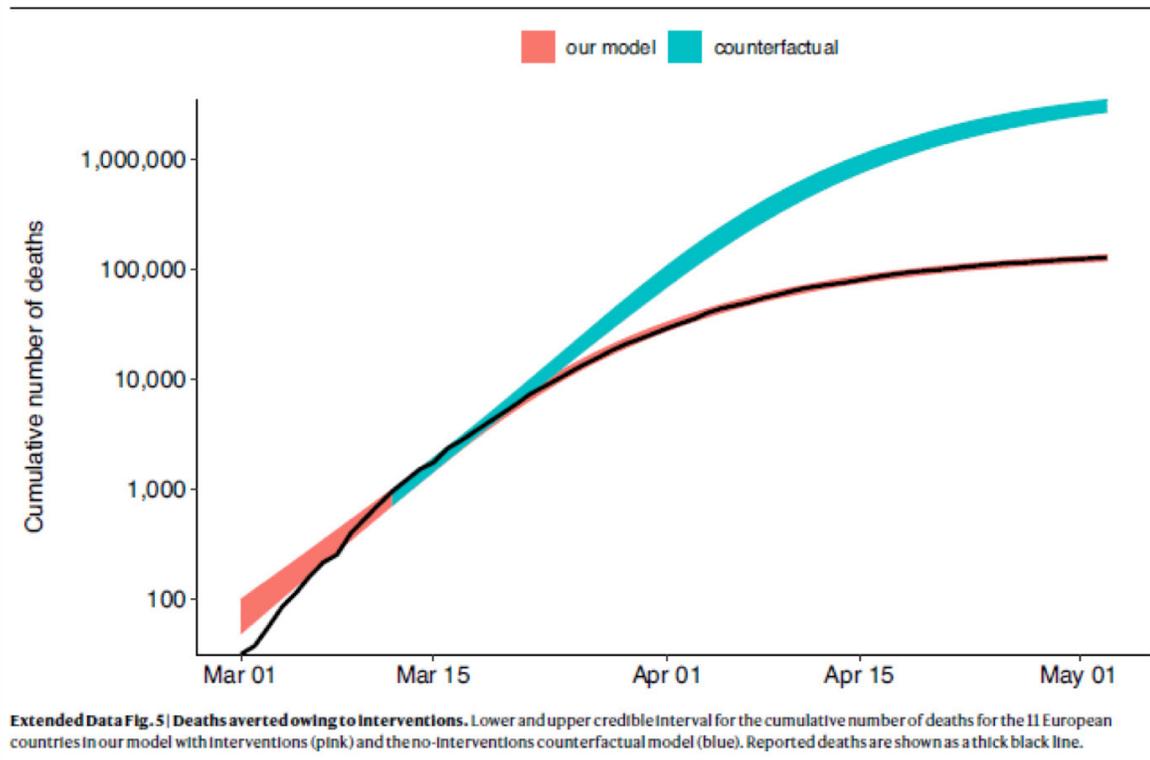
18. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2019: Data Booklet*, ST/ESA/SER/A/424. (United Nations, 2019).

19. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).

20. Goldstein, E. et al. Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proc. Natl Acad. Sci. USA* **106**, 21825–21829 (2009).

**Acknowledgements** S.B. acknowledges the NIHR BRC Imperial College NHS Trust Infection and COVID themes, the Academy of Medical Sciences Springboard award and the Bill and Melinda Gates Foundation. L.C.O. acknowledges funding from a UK Royal Society fellowship. Initial research on covariates in Supplementary Table 2 was crowdsourced; we thank a number of people across the world for help with this. This work was supported by Centre funding from the UK Medical Research Council under a concordat with the UK Department for International Development, the NIHR Health Protection Research Unit in Modelling Methodology and Community Jameel. We thank F. Valka for creating our website, and A. Gelman and the Stan

Article



**EPIDEMAP : UN OUTIL DE SIMULATION SPATIALISÉE D'ÉPIDÉMIE**

<https://www.data.gouv.fr/fr/reuses/modelisation-de-levolution-des-epidemies-covid-19/>

Publié le 17 avril 2020 par Epidemap (outil de modélisation développé au CEA). Le croisement de 3 modèles a permis de simuler la propagation du CoViD-19, habitant par habitant, dans toute la France. L'objectif est de pouvoir simuler l'impact de différents scénarios de déconfinement sur l'impact de l'épidémie sur la population, avec une granularité géographique à l'échelle des bâtiments, et des tranches d'âge par intervalle de 10 ans.

- Un premier modèle utilise les données opendata pour la répartition des habitants, bâtiment par bâtiment ;
- Un second modèle permet de simuler le déplacement de chaque personne ;
- Un troisième modèle d'épidémiologie permet de simuler la propagation du virus entre chaque individu, jour après jour.

Toutes les informations et résultats se trouvent sur <http://www.epidemap.com>

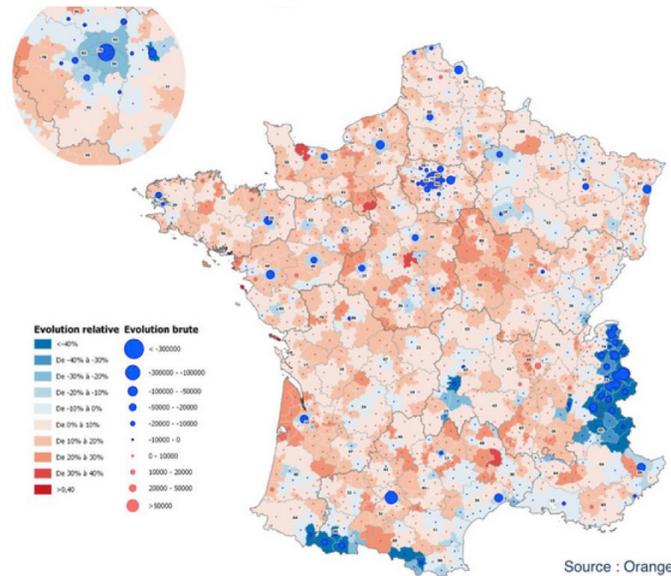


## ANNEXE TECHNIQUE ET BIBLIOGRAPHIQUE :

### 3. La question des données



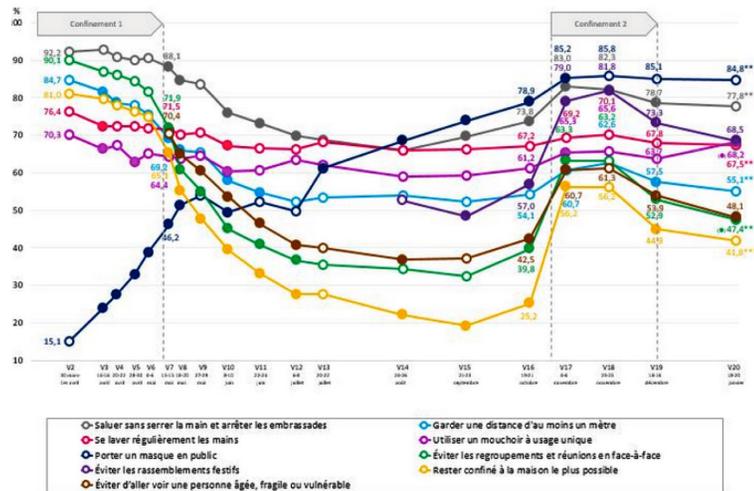
➤ *Suivi des positions (opérateurs mobiles : Orange, SFR, ...) :*



Évolution du nombre de téléphones mobiles par zone géographique entre le 13 et le 20 mars 2020

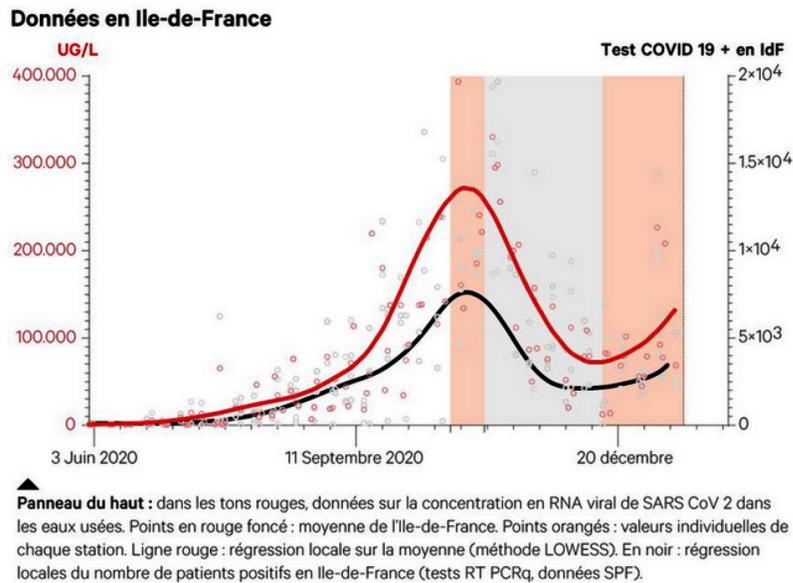
➤ *Indicateurs de suivi des comportements et psychologique (cf Santé publique)*

<https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19>

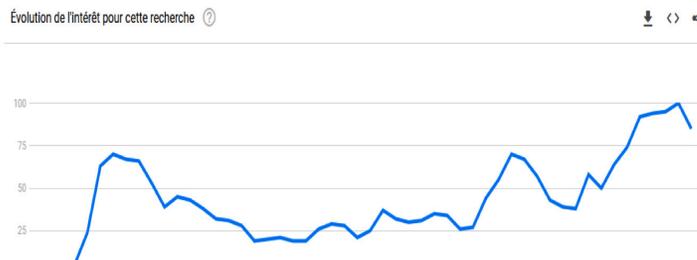


➤ *Indicateurs indirects de propagation du virus : suivi des eaux usées*

<https://www.lesechos.fr/industrie-services/energie-environnement/covid-les-eaux-usees-revelent-une-inquietante-montee-du-virus-1281066> )



➤ *Indicateurs de recherche de mots-clés dans les moteurs de recherche, qui se corrèle assez bien avec le taux de décès*



Et d'autres indicateurs plus exotiques comme le taux d'occupation des parkings des hôpitaux mesuré par satellite.

Nous n'avons pas pu obtenir de données économiques de suivi au quotidien, mais on aurait pu collecter des indicateurs comme les mouvements des ports, les mouvements d'avions, la collecte de TVA, le niveau des transactions bancaires, etc., qui donnent des informations économiques avancées sur les l'état du PIB.

## LA COLLECTE

Le coût de collecte et d'organisation de ces données et indicateurs est très variable, résultant de traitements habituels pour des entreprises de gestion de données (opérateurs, Google...), à des extraits spécifiques (données réelles de transport public), ou collectées semi-manuellement (données des hôpitaux et praticiens de santé). Si beaucoup de données agrégées ont été cédées gratuitement aux pouvoirs publics, on peut aussi facilement imaginer que le coût de mettre des informations à disposition des pouvoirs publics soit compensé financièrement.

La collecte de ces informations pose des problèmes techniques d'anonymisation (géolocalisation par les opérateurs et les GAFA), voire de droits (peut-on utiliser les données d'un individu chez un opérateur pour dénoncer un déconfinement abusif ?), voire des problèmes purement techniques (le suivi par bornage est plus complexe que par simple géolocalisation). Dans le cadre du RGPD, se pose la question du consentement de l'utilisateur à un traitement qui n'est pas prévu : l'opérateur peut-il publier des données ? L'une des questions soulevées est celle du stockage de ces données. « *Cela demande des ajustements réglementaires et un accord de la CNIL. Il faudrait pouvoir garder des données sur une durée de temps longue, or actuellement nous devons les supprimer au bout d'un an* », a reconnu Stéphane Richard, le PDG d'Orange, au [Figaro](#).

La durée et le retard de collecte est aussi une difficulté qui a été rencontrée : la collecte manuelle des décès COVID faiblissait considérablement le week-end pour bondir le lundi. Autres soucis de saisie pour les données des EHPAD, qui ont été agrégées plusieurs fois sur des périodes de plusieurs jours vers un seul jour de déclaration, sans oublier que le département des Bouches-du-Rhône n'a pas fourni de données pendant plusieurs semaines en avril.

De même, les enquêtes psychologiques de terrain sont difficiles à suivre au quotidien et à mesurer.

## VALIDATION ET INTERPRÉTATION

La qualité des données est un vrai sujet. Changer la méthode de comptage en cours de route est une vraie question ! On l'a observée dans le décompte des cas COVID :

- en mars/avril les cas décomptés sont essentiellement très proches des arrivées en hôpital ;
- ensuite la mesure prise en compte est le nombre de tests PCR positifs, mais qui dépendait essentiellement du nombre de tests réalisés : le nombre de cas augment juste en proportion du nombre de tests ;
- puis le nombre de cas est calculé comme le ratio (nombre de cas positifs/nombre de tests)
- enfin ce ratio est modifié le 8 décembre, pour 'prendre en compte' les cas de test multiples de personnes se faisant tester plusieurs fois.

Évidemment, ces variations de méthode introduisent une forte discontinuité d'interprétation, mais ne sont pas recensés explicitement dans des bases différentes. Donnons deux exemples sur la méthodologie de comptage qui n'est pas si simple à définir :

- Le débat de savoir si les décès ont tous été PAR le COVID ou AVEC le COVID reste entier à ce stade. Le nombre de décès avec des comorbidités élevées est tellement important que la mortalité ajoutée n'est pas toujours évidente à définir : s'il n'y a pas de surmortalité comme en certaines périodes, mais qu'il y a des décès catalogués COVID, que doit-on en conclure ?
- La réalisation de campagnes de tests PCR très importantes et coûteuses a permis de remonter des séries de comptage de positivité. Cependant, le nombre de cycles utilisé n'est pas reporté dans les statistiques, alors qu'il a été établi que le taux d'erreur grandit avec le nombre de cycles. On aurait ainsi souhaité disposer de résultats quotidiens sous forme de couples (nombre de cycles PCR, nombre de cas détectés positifs), au lieu de leur simple agrégation.

Certaines données (eaux usées, mobilités) sont entachées de grandes marges d'erreurs, et ne peuvent être utilisées au mieux que comme indicateur en variation, car elles ne peuvent être corrélées de façon fiable au niveau de santé de la population.

Malgré tous leurs défauts, les données ont permis néanmoins de découvrir des informations intéressantes sur les mécanismes de propagation de la maladie. Le décalage entre les pics des différentes courbes renseigne sur les durées de propagation, d'infection et la létalité.

## L'IMPORTANCE DU TEMPS

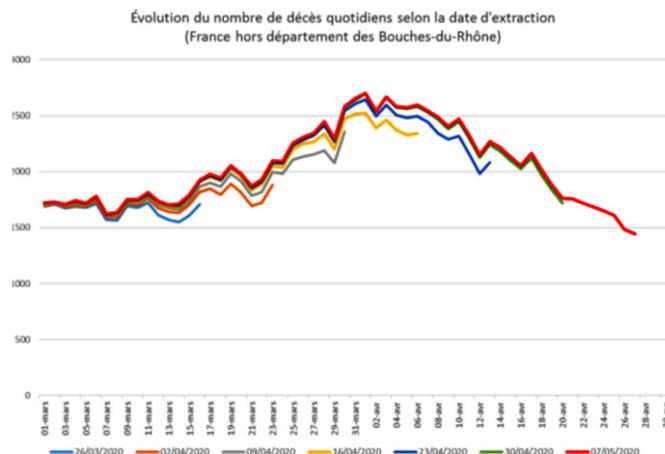
La constitution de séries temporelles est indispensable pour obtenir des prévisions. Encore faut-il les produire ... à temps !

Plusieurs points ont été notés dans cette période COVID :

- Les données sont disponibles tardivement ;
- Les données sont susceptibles d'évolutions au cours du temps.

Un exemple simple est fourni par les statistiques de décès en France de l'Insee : la remontée des décès par les mairies prend du temps, qui se compte en jours pour 90 % des cas, mais prend des mois voire des années pour être comptabilisée très précisément. La courbe publiée par l'Insee au jour le jour montre ainsi une sous-estimation d'au moins 15 % par rapport à un chiffre plus juste qui arrive avec six semaines de délai.

La disponibilité d'indicateurs avancés est un point clé. Si on n'en dispose pas, seule la modélisation fine des erreurs systématiques de mesure permet d'approcher des chiffres réalistes en temps réel.



## CONCLUSIONS SUR LES DONNÉES

Un épisode de crise aussi complexe que le COVID a amené à utiliser des données de nature très variables. Il faut reconnaître qu'un effort important et inédit a été fait pour partager les données disponibles. Il apparaît toutefois que la documentation des différentes méthodes de collecte de données doit être encore renforcée. Devant cette variété de difficultés, il semble que les méthodologies de collecte de données sont à mieux documenter afin de pouvoir adapter les algorithmes et interprétations, notamment pour disposer d'indicateurs avancés, pour corréler différentes sources et conforter des données existantes.

Enfin il ne faut pas oublier l'interprétation visuelle des courbes de données.



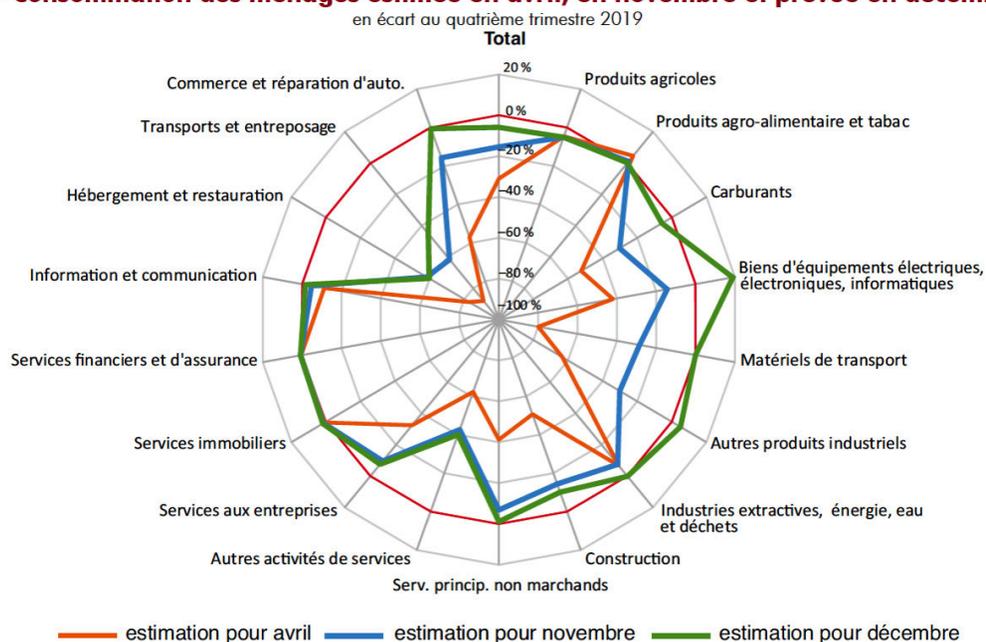
## ANNEXE TECHNIQUE ET BIBLIOGRAPHIQUE :

### 4. Aspects économiques, sociaux et sociétaux

## IMPACT SUR L'ÉCONOMIE

Concernant la conjoncture économique (15/12/2020) pour la France, voir la note [https://www.insee.fr/fr/statistiques/fichier/4997475/note-conj\\_151220.pdf](https://www.insee.fr/fr/statistiques/fichier/4997475/note-conj_151220.pdf)

### 6 - Consommation des ménages estimée en avril, en novembre et prévue en décembre



Lecture : la perte d'activité dans l'hébergement et la restauration est estimée à -60 % en décembre 2020 contre -83 % estimée en avril (en écart au quatrième trimestre 2019).

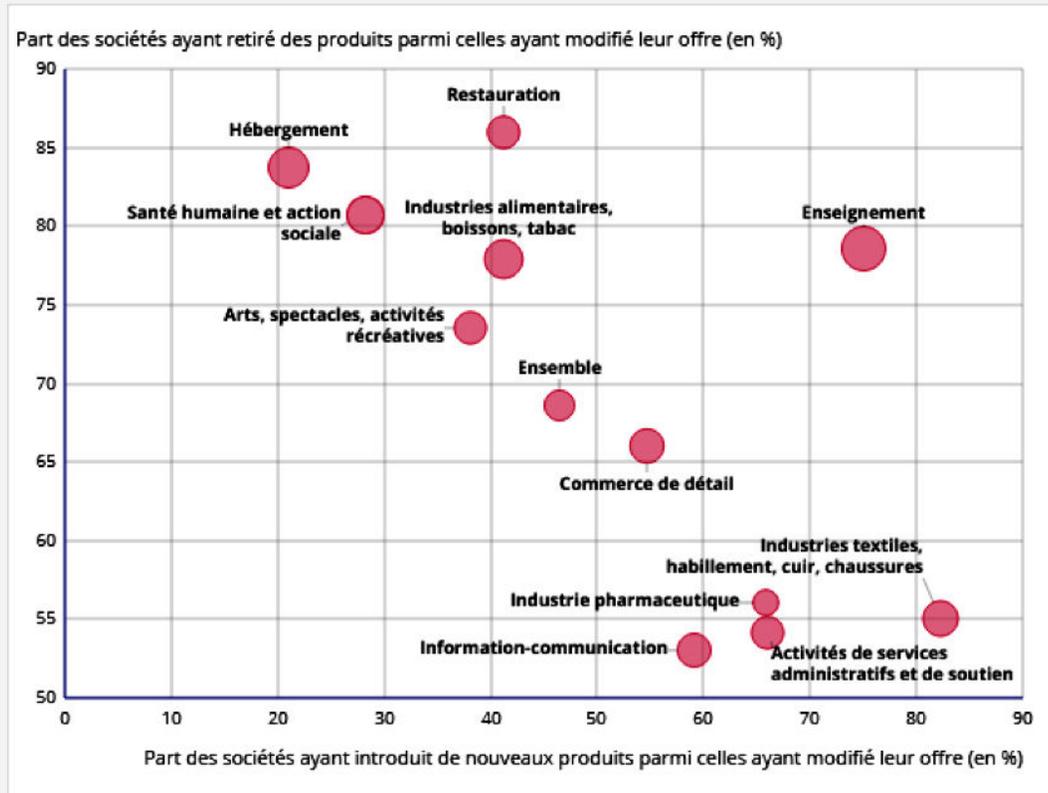
Source : calculs Insee à partir de sources diverses

Le diagramme ci-dessus est extrait de cette note il illustre l'effet de la pandémie sur la consommation des ménages.

Concernant l'impact sur les sociétés en France, voir l'article de l'INSEE:

<https://www.insee.fr/fr/statistiques/4994488>

**Figure 5 – Modification de l'offre, retrait ou introduction de nouveaux produits ou services pendant le premier confinement**



Note : la taille des ronds représente la proportion de sociétés ayant modifié leur offre.

Lecture : dans l'industrie pharmaceutique, 14 % des sociétés ont modifié leur offre pendant le premier confinement (taille du rond). Parmi celles-ci, 66 % ont proposé de nouveaux produits ou services, 56 % en ont délaissé.

Champ : sociétés de 10 salariés ou plus, sauf pour le commerce de détail (sociétés de 5 salariés ou plus).

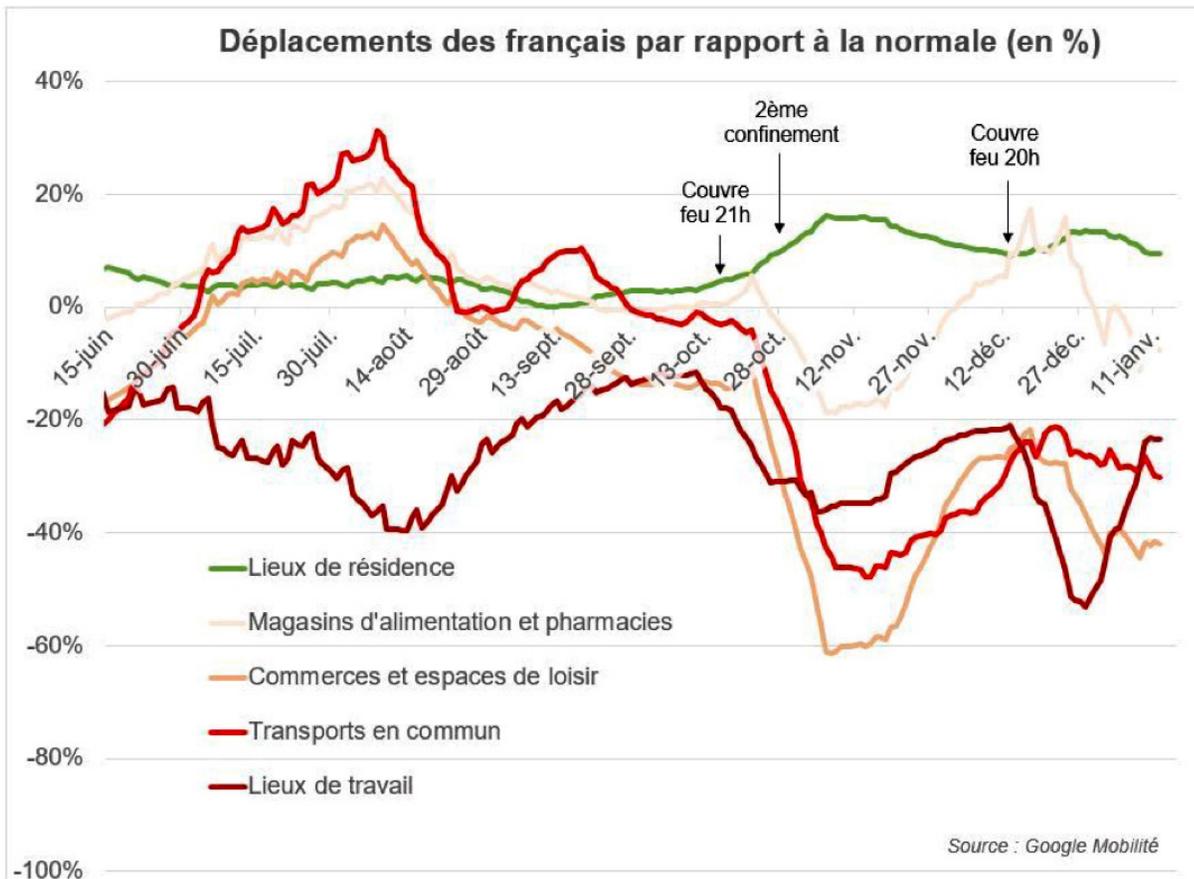
Source : Insee, enquête Impact de la crise sanitaire sur l'organisation et l'activité des sociétés, 2020.

Le schéma ci-dessus illustre la profondeur de l'impact sur les sociétés et les effets de réorganisation qui s'en sont suivis. Cette étude d'impact est accompagnée de données publiées [https://www.insee.fr/fr/statistiques/fichier/4994488/IP1830\\_donnees.xlsx](https://www.insee.fr/fr/statistiques/fichier/4994488/IP1830_donnees.xlsx).

## LES MOBILITÉS

- <https://www.google.com/covid19/mobility/> Ce site donne les impacts sur les mobilités dans les régions du monde

*Le graphique qui suit en est extraite, à propos de laquelle on peut formuler les observations suivantes. Il y a eu plusieurs démarches pour montrer que la courbe de réduction des entrées en réanimation avait diminué juste au moment du second confinement, donc que ce nouveau confinement n'avait pas d'effet, puisqu'il y aurait dû y avoir un décalage de 7 à 10 jours entre cause et effet. En fait ces courbes montrent bien que la posture de confinement (moins de déplacements, de transports en commun, etc.) a démarré au moins 8 jours avant l'annonce du confinement. Explication probable : la menace a un effet dissuasif. Cette observation plaide en faveur d'une démarche de la modélisation psychologique.*



- [https://en.wikipedia.org/wiki/Social\\_impact\\_of\\_the\\_COVID-19\\_pandemic](https://en.wikipedia.org/wiki/Social_impact_of_the_COVID-19_pandemic)  
[https://en.wikipedia.org/wiki/Social\\_impact\\_of\\_the\\_COVID-19\\_pandemic\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/Social_impact_of_the_COVID-19_pandemic_in_the_United_States)
- <https://www.tandfonline.com/doi/full/10.1080/02185385.2020.1819396>  
*The first COVID-19 case in Taiwan was confirmed on 21 January 2020. Compared to the global case fatality rate, the people in Taiwan have been affected moderately by this disease, from a medical point of view. However, when looking back to their life experiences over the past six months, one must pay attention to the social dimensions of this pandemic. In this article, we examine four key social issues that can provide some insights into the social work imagination, after this worldwide health crisis is over.*
- [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30800-8/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30800-8/fulltext)  
*China's successful control of COVID-19. While the world is struggling to control COVID-19, China has managed to control the pandemic rapidly and effectively. How was that possible? Talha Burki reports.*

## INDUSTRIE ET LOGISTIQUE

- [https://www.ifc.org/wps/wcm/connect/industry\\_ext\\_content/ifc\\_external\\_corporate\\_site/infrastructure/resources/the+impact+of+covid-19+on+logistics](https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/infrastructure/resources/the+impact+of+covid-19+on+logistics)  
IFC: International Finance Corporation
- <https://www.nature.com/articles/d41586-020-03134-2> Nature, vaccine logistics  
Title of article: COVID vaccination logistics: five steps to take now.

## LE CAS DE LA CORÉE DU SUD

### *SUR LA PANDÉMIE*

- <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-020-01791-8>  
Bon article sur l'épidémie en Corée du Sud graphiques intéressants.

### *ASPECTS SOCIÉTAUX ET DIVERS*

- <http://www.korea.net/NewsFocus/Society/view?articleId=185127>  
*Mme Ki Moran : Ki also chairs the COVID-19 policy committee of the Korean Society for Preventive Medicine*
- [https://www.eurekalert.org/pub\\_releases/2020-08/uocd-wtr081020.php](https://www.eurekalert.org/pub_releases/2020-08/uocd-wtr081020.php)
- Emerging COVID-19 success story: South Korea learned the lessons of MERS: <https://ourworldindata.org/covid-exemplar-south-korea>



La crise du Covid-19 a rapidement dépassé le domaine du pur médical pour impacter l'ensemble de l'économie et des activités françaises. Le système « France » a été soumis à une série de chocs majeurs successifs (santé, confinements, économie, psychologie...), et a été gravement perturbé.

La prise de décisions nouvelles s'est imposé au gouvernement et aussi à tous les acteurs de l'économie : comprendre le présent, anticiper le futur, comprendre les leviers d'actions possibles, étudier des scénarii sont devenus des éléments clés pour chacun à son échelle.

Pour améliorer les décisions et éviter les erreurs majeures, le rapport préconise le développement d'une plate-forme numérique permettant d'agrèger les données et moyens du terrain et de supporter de nouveaux modèles de prévision et simulation variés et renouvelés (recherche opérationnelle, théorie des jeux, fouille de données et de textes, etc.), et surtout rapidement évolutive. Les approches pluridisciplinaires en matière de modélisation sont indispensables et à encourager, y compris autour de l'épidémiologie, et nos forces en mathématique et informatique, tant académiques qu'industrielles, doivent être mobilisées dans ce but

Académie des technologies  
Le Ponant – Bâtiment A  
19, rue Leblanc  
75015 PARIS  
+33(0)1 53 85 44 44  
secretariat@academie-technologies.fr  
www.academie-technologies.fr

©2020 Académie des technologies  
ISBN : 979-10-97579-29-6

