



Des données pour évaluer...

La mobilisation des données pour le suivi et l'évaluation des politiques publiques.

Séminaire commun Cour des comptes et Société française de l'évaluation

Jeudi 8 octobre 2015, 14 h – 17 h 15.

En Grand'chambre, Cour des comptes - 13 rue Cambon - Paris 1^{er}.

Les raisons de progresser ensemble sur le thème des données pour évaluer.

Les ingrédients cruciaux d'une évaluation de qualité de politiques publiques et de programmes sont aujourd'hui dans les sources et les données (« data »).

Les données sur les politiques publiques et les programmes et sur l'environnement socio-économique de ces politiques et programmes sont des données de plus en plus ouvertes. Mais, ouverture ne signifie pas facilité d'utilisation. Donc, paradoxalement, l'évaluation manque encore souvent de moyens en données et connaissances mobilisables et à la hauteur des mandats des cahiers des charges des évaluations commanditées.

Plusieurs questions sont de ce fait à la recherche de leurs réponses : comment obtenir une meilleure adaptation des programmes d'enquête et de statistique publique, au-delà des besoins d'observation, pour mieux comprendre les effets économiques, sociaux, environnementaux, territoriaux des politiques publiques et des programmes ?

Pourquoi les données de gestion publique des programmes dans les différents domaines sectoriels (administrations, collectivités, délégataires de service public) ou par rapport aux différents publics administrés ou aidés sont-elles insuffisamment mobilisées alors que, rendues interopérables et couplées aux données de contexte ou de grandes enquêtes de la statistique publique, elles seraient utiles pour l'évaluation avant, pendant et après ?

Idéalement, les évaluateurs qui doivent produire de la connaissance sur les résultats (efficacité) et la valeur (efficience) des politiques publiques et des programmes devraient pouvoir mobiliser et disposer de suffisamment d'informations contenues dans des données accessibles, protégées, rendues interopérables sur des plateformes neutres et ceci quelle que soit la source. Sur ces plateformes, ils pourraient, au gré de leurs interrogations, disposer d'informations faciles à utiliser (fiables, les plus récentes possibles, multi-sources, brutes ou retravaillées) pour développer la connaissance, suivre et évaluer ?

Ainsi serait amélioré le cycle « données-informations- connaissances-actions ».

Malheureusement, l'organisation industrielle de la filière de la donnée et la place du service public de la donnée n'en sont encore qu'à des balbutiements.

La Cour des Comptes et la Société française d'Évaluation, préoccupées par la qualité du travail de suivi évaluatif des programmes et d'Évaluation des politiques publiques, ont réuni pour la première fois différents producteurs de données et différents utilisateurs pour commencer à répondre à la question jugée pertinente suivante : *« Quelles recommandations peut-on aujourd'hui faire en France pour mieux produire, mobiliser, rendre plus riches et faciliter les différentes utilisations des sources des données publiques et de gestion publique ou de délégataires de service public à des fins d'extraire la valeur ou l'intelligence (« smart data ») à des fins de suivi et d'évaluation ? »*

Maurice Baslé, président d'honneur de la société française d'évaluation.
Maurice.basle(a)univ-rennes1.fr

Gilles-Pierre Lévy, président de Chambre à la Cour des comptes.
Gplevy(a)ccomptes.fr

Introduction générale.

Didier Migaud. Premier président Cour des comptes

En pleine expansion sur le plan international, l'évaluation des politiques publiques tend à se développer en France au sein de nombreuses institutions. C'est le cas à la Cour des comptes, qui s'est vue confier depuis la révision constitutionnelle du 23 juillet 2008 un nouveau rôle en la matière, celui d'assister le Parlement dans sa mission d'évaluation des politiques publiques. « L'évaluation par la donnée est un prolongement de notre activité de contrôle par la pièce », a souligné Didier Migaud, premier président de la Cour.

Introduction à l'innovation des données ouvertes.

Henri Verdier, directeur interministériel du numérique et du système d'information et de communication de l'État.



Il n'est pas exagéré de dire que la révolution numérique, qui bouleverse la plupart des activités humaines, est en passe de devenir une révolution de la donnée, note Henri Verdier (photo), directeur interministériel du numérique et du système d'information et de communication de l'Etat et administrateur général des données, au sein du secrétariat général pour la modernisation de l'action publique (SGMAP).

Plusieurs paramètres sous-tendent cette révolution :

- la production de quantités quasi infinies de données, par d'innombrables sources : capteurs bon marchés, nouveaux comportements et notamment dans le web social, traces d'utilisation de nombreux dispositifs, et surtout l'informatisation de nombreux processus. Avec cette révolution, ce n'est pas seulement la quantité des données qui change, mais lus encore leur nature. Là où la plupart des données étaient construites par la science et pour les besoins de la connaissance, elles sont produites par des activités spontanées, ou des dispositifs de gestion, sans qu'on sache encore très bien tout ce qu'elles recèlent ;
- la facilitation considérable de leur hébergement, de leur transfert et de leur traitement, notamment grâce de nouvelles technologies, comme le big data, le cloud, traitement temps réel, et la baisse corrélative des coûts de ces traitements, dont une conséquence imprévue est la démultiplication des organisations et des personnes capables de traiter ces données ;
- et enfin – et c'est peut-être le plus important – la diversification des questions posées à ces données et des stratégies fondées sur ces données. On ne se contente plus de les utiliser pour décrire le réel, ni même de le prédire (analytique prédictive), mais on leur demande aussi de prescrire des actions (boucles de rétrocontrôle, systèmes de décision temps réel, recherche de la « prochaine action à mener », etc.)

Cette révolution ouvre tout un champ des possibles à l'évaluation.

D'une part, et ce n'est pas toujours le plus facile, elle doit, comme toutes les sciences, apprendre à sortir d'un monde où les données étaient rares et chères – mais travaillées et fiables – à un monde où elles sont abondantes et quasi gratuites, mais produites pour des fins très hétérogènes. Cette conversion de l'attitude du chercheur n'est pas la moindre des transformations pour les scientifiques, dans aucun domaine.

D'autre part, elle ouvre sur de nouvelles logiques de l'évaluation : évaluation en temps réel, évaluation collaborative, etc.

Enfin, elle nous conduira probablement à changer les demandes mêmes qui seront formulées aux évaluateurs, qui devront apprendre progressivement à passer d'une logique de description fiable du réel à une logique de conduite de l'action, et descendre ainsi de plus en plus profondément au cœur des processus et en particulier des politiques publiques.

Lire davantage : <http://www.modernisation.gouv.fr/laction-publique-se-transforme/en-évaluant-ses-politiques-publiques/des-donnees-pour-evaluer>

L'Age de la multitude. *Entreprendre et gouverner après la révolution numérique*, de Nicolas Colin et Henri Verdier. Armand Colin, 22,90 euros, 285 pages.

Chapitre un. Les mobilisations des données pour le suivi et l'évaluation des politiques publiques dans le domaine de la santé.



Gilles Pierre Levy, président de chambre à la Cour des comptes.

Le domaine de la santé est tout naturellement au carrefour des politiques publiques et de la manipulation d'un grand volume de données.

Politiques publiques tout d'abord dans la mesure où c'est un des secteurs où, même dans les pays les plus libéraux, l'intervention de la Puissance publique est jugée nécessaire, au minimum pour assurer la qualité des soins et des médicaments –cf. le rôle de la « Food and Drug Administration » aux Etats-Unis –, et plus souvent pour garantir soit la fourniture directe, soit le financement de tout ou partie des actions en faveur de la santé.

Manipulation d'un grand volume de données, ensuite, à la fois du fait de l'ancienneté des études épidémiologiques, de la gestion des données de financement des soins par les organismes de Sécurité sociale, mais aussi par les assureurs privés ou les réseaux de soins quand il en existe. Cette masse de données est aujourd'hui considérablement accrue par les informations qui circulent sur les réseaux sociaux, par les informations que recueillent Google, Apple, Amazon, etc... sur les forums qui concernent directement ou indirectement la santé, sur le suivi des cardiaques, sur les achats de médicaments non remboursés, sur les achats alimentaires, etc... La liste est longue et surtout difficile à limiter tant des sujets connexes peuvent avoir un impact sur la santé et partant sur la définition et l'évaluation des politiques en cause.

Mais l'usage de ces données multiples pour évaluer les politiques de santé soulève de nombreuses questions : adaptation de données de remboursement à l'appréciation de politiques par nature pas seulement financières ; gratuité des données fournies par certains –les administrations, par exemple –et revendues par d'autres –tels que certains sites privés; localisation des bases qui centralisent ces données dans d'autres pays, tels que les Etats – Unis, où sont installés nombre de sites; protection de la vie privée ; coût de l'exploitation des chiffres ;etc.

Le potentiel pour une meilleure évaluation des actions publiques est énorme. Sa maîtrise reste à renforcer.

1.1.La mobilisation des données en santé : état des lieux et enjeux.

Franck Von Lennep, directeur de la DREES (Direction de la recherche, des études, de l'évaluation et des statistiques).

L'évaluation en santé, qu'elle relève des administrations ou des chercheurs, est en France plus développée qu'on ne le croit mais moins qu'on ne pourrait le souhaiter. L'enrichissement de bases de données accessibles est essentiel pour favoriser l'essor de l'évaluation.

Quelques-uns des principaux enjeux relatifs aux données de santé seront ici rapidement abordés.

L'enjeu général de l'open data.

Cet enjeu concerne l'ensemble des producteurs de données de santé. Il s'agit de multiplier les données accessibles à tous et réutilisables sans restriction. Ces données peuvent ainsi être librement rapprochées d'autres sources.

En la matière, il convient de ne pas sous-estimer l'existant (voir par exemple les sites éco-santé, score-santé, ameli, data.drees, atih, ...).

Les principaux acteurs institutionnels du champ de la santé ont une politique volontariste consistant à démultiplier le nombre de ces données en open data.

Lorsque ces jeux de données sont produits à partir de données individuelles, un enjeu important est de s'assurer de leur parfaite anonymisation (ce qui signifie que le risque de ré-identification d'une personne doit être nul) avant leur mise en ligne.

L'enjeu des bases de données médico-administratives (Sniiram-PMSI, futur système National des Données de Santé).

Les données médico-administratives du Sniiram-PMSI (données de remboursement en ville et séjours hospitaliers) sont d'une grande richesse et permettent de mener des travaux d'évaluation approfondis du système de santé. Il est impératif de poursuivre le mouvement engagé depuis plusieurs années et de faciliter et développer le recours à ces données : c'est l'objet de l'article 47 du projet de loi de modernisation de notre système de santé. En particulier, l'article 47, lorsque la loi sera définitivement adoptée, permettra d'élargir les accès permanents pour les institutions exerçant une mission de service public : ministère de la santé, caisses de sécurité sociale, agences nationales, Agences régionales de santé, qui pourront ainsi accroître les travaux d'études et d'évaluation. Il facilitera également les appariements de données en santé en rendant possibles ces appariements à partir du numéro de sécurité sociale (NIR) avec l'autorisation de la CNIL mais sans passer par un décret un Conseil d'Etat comme actuellement.

L'enjeu des enquêtes et cohortes.

Les bases médico-administratives présentent deux avantages immenses par rapport aux enquêtes : leur exhaustivité (permettant notamment les études régionalisées) et leur coût (puisqu'elles sont issues de données de gestion).

Mais le futur système national des données de santé (SNDS) ne couvre pas tout le champ des données nécessaires à l'évaluation (situation sociale des patients, facteurs de risques et comportements, diagnostics, résultats d'examens de biologie ou d'imagerie...).

Demeure donc un enjeu majeur autour des enquêtes et cohortes et des appariements entre ces enquêtes et les données médico-administratives (par exemple, la cohorte Constance de l'INSERM, l'enquête CARE de la DREES sur les personnes dépendantes...).

Il faut également souligner l'apport, pour l'évaluation, des appariements entre bases de données administratives. On peut citer en exemple la base Hygie de l'IRDES qui apparie données de l'assurance maladie et données de carrière recueillies par la CNAV ; ou encore le projet ambitieux porté par la DREES et l'INSEE d'appariement entre le SNDS et l'échantillon démographique permanent de l'INSEE.

L'enjeu des compétences et des financements.

La mise à disposition de données médico-administratives est d'autant plus coûteuse que le nombre de demandeurs s'accroît. La question des moyens à l'assurance maladie pour développer la mise à disposition de ces données est ainsi centrale.

Il ne faut pas ignorer par ailleurs la question des compétences dans les institutions publiques pour développer l'expertise sur ces bases de données, qui sont d'un maniement complexe. Citons les exemples de la DREES, qui développe un pôle de compétences en son sein, et des ARS, qui devront investir sur ce sujet lorsqu'elles disposeront d'accès élargis aux données avec l'application de l'article 47.

Enfin, dès lors que les bases de données médico-administratives ne suffisent pas à satisfaire par elles-mêmes la totalité des besoins d'évaluation, un enjeu fondamental pour la statistique publique et les chercheurs est de disposer de moyens financiers suffisants pour continuer de financer des enquêtes et des cohortes.

En conclusion, au-delà des données médico-administratives et de leurs appariements avec des données d'enquêtes ou avec d'autres données administratives, on voit poindre pour l'avenir des interrogations sur des données « privées » qui pourraient enrichir les analyses. On pourrait alors réfléchir à rapprocher les données habituelles de la sphère santé avec des données externes. Un exemple pourrait ainsi consister à utiliser des données individuelles sur les trajets (téléphonie mobile, transports publics, ultérieurement voitures connectées) pour mieux travailler sur la répartition spatiale de l'offre de soins.

Ainsi, au-delà des enjeux liés aux données produites par les administrations, les chercheurs et les administrations s'interrogent de manière croissante sur les apports possibles des données recueillies par des acteurs privés, et sur les utilisations envisageables de l'ensemble de ces données (publiques et privées). Le Ministère de la santé a lancé une vaste réflexion sur ce sujet du big data, organisée autour de quatre groupes de travail pilotés par des binômes public/privé, qui remettront leurs réflexions au début de l'année 2016.

1.2. Les besoins des évaluations de plans de Santé Publique.

Franck Chauvin, membre du collège du Haut Conseil de la Santé publique.

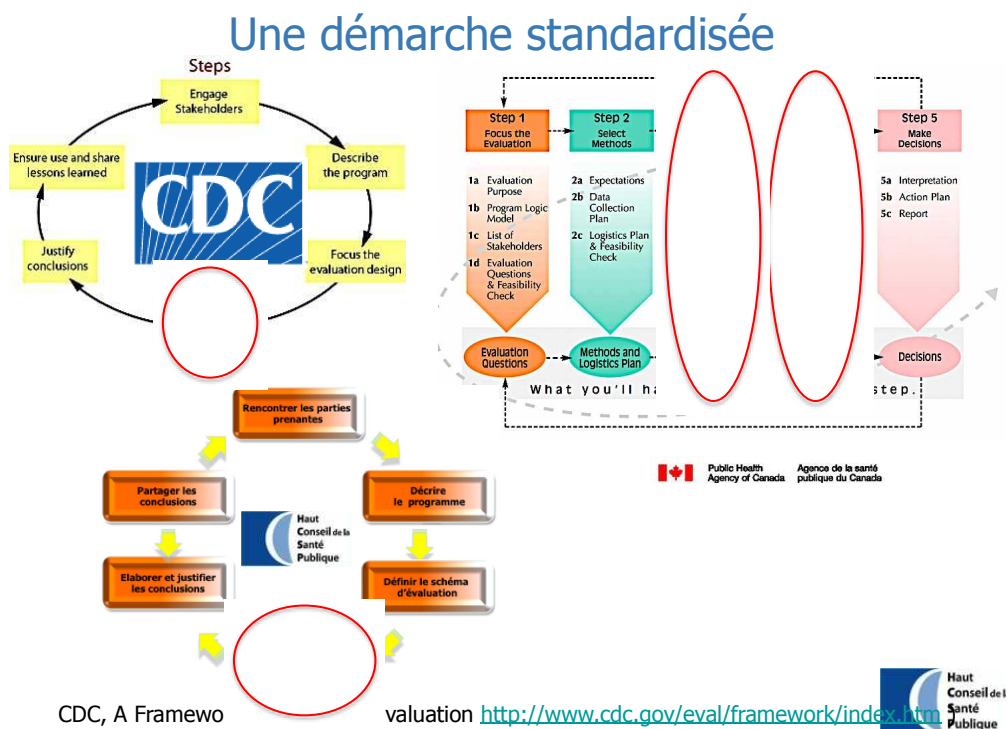
Le Haut Conseil de la Santé Publique et l'évaluation des politiques publiques de santé

Le Haut Conseil de la Santé Publique (HCSP) a été créé par la loi de Santé publique de 2004 et installé en 2007. Il a dans ses missions l'aide à la gestion des risques sanitaires et l'évaluation des politiques publiques de santé et notamment l'évaluation des plans de santé publique. Une commission spécifique, « Evaluation-Stratégie-Prospective » est plus spécialement chargée de cet aspect et évalue en moyenne 4 plans de santé publique par an, sur saisine du ministère des affaires sociales, de la santé et du droit des femmes.

Parallèlement, la ministre de la santé a saisi le HCSP en février 2014 pour élaborer un tableau de bord synthétique de l'état de santé de la population et du fonctionnement du système de santé, destiné au pilotage interministériel de la Stratégie nationale de santé et à une communication annuelle auprès du grand public. Enfin, le HCSP peut être saisi en amont de la mise en œuvre d'un plan de santé publique pour émettre un avis sur la construction d'un plan et son évaluabilité.

Ainsi ses missions l'amènent à être un utilisateur de données produites par différents organismes.

Dans son rôle d'évaluateur, le HCSP consacre une part importante de ses ressources et du temps de l'évaluation à collecter les données nécessaires, aussi bien quantitatives, lorsqu'elles existent, que qualitatives lors d'entretiens ou d'enquêtes ad hoc. En cela, il applique la même méthodologie que celle mise en œuvre par les institutions similaires des autres pays (« Centre for Disease control » américain, agence canadienne ...).



Les données : une nécessité pour l'évaluation.

L'évaluateur souhaite toujours disposer de plus de données pour évaluer les effets de politiques publiques, qu'il s'agisse de l'impact, de l'efficacité ou à un moindre degré du suivi de la mise en œuvre de mesures ou de programmes.

Cependant, l'évaluation des politiques publiques de santé passe, autant que possible, par l'accès à des données agrégées sous la forme d'indicateurs. Ces indicateurs issus de systèmes d'information multiples sont malheureusement peu souvent disponibles et leur évolution difficile à analyser au regard de la ou des questions évaluatives.

De ce point de vue, l'accès à un nombre de plus en plus important de données concernant la santé à la fois représente une opportunité.

L'accès aux données de santé : une opportunité à tempérer probablement.

Une opportunité parce que l'évidence de l'efficacité d'une mesure ou d'un plan est issue de la confrontation de points de vue différents, comme c'est le cas pour l'évidence scientifique. L'accès à des sources différentes devrait renforcer la pertinence des conclusions de l'évaluation.

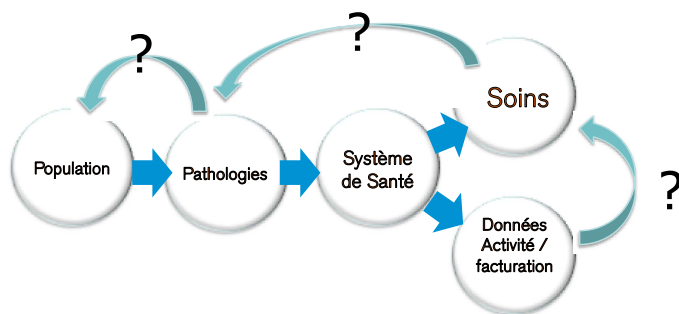
Des données, issues de systèmes d'information jusque-là fermés, sont maintenant potentiellement accessibles. C'est le cas des données issues du SNIRAM, système d'information de l'assurance maladie, des données d'activité des hôpitaux issues du PMSI (programme de médicalisation du système d'information), des données issues du DMP (dossier médical partagé) etc... Ces bases de données permettront en principe de compléter la description de l'état de santé d'une population ou de sous-populations.

Il faut toutefois noter deux biais principaux inhérents à l'utilisation de ces bases :

- la description de l'état de santé se fera à travers des systèmes recueillant des données de maladie. De ce fait, elles ne permettront de ne répondre qu'à des questions concernant la santé définie comme absence de maladie. Cette approche opérationnelle est évidemment très réductive et des champs de la santé, comme la santé perçue, sont exclus de cette description.
- Ces systèmes sont pour les deux premiers des systèmes de facturation et non de description et ne permettent donc de décrire un état de santé qu'à travers la consommation de soins.

Ainsi l'utilisation de ces bases pose la question du modèle explicatif que l'on veut utiliser : peut-on décrire les soins par la « facturation des soins » ? Peut-on décrire les pathologies par les données de soins ? Les pathologies rendent-elles compte de la santé d'une population ? (cf. schéma)

Système de santé et données



Les questions méthodologiques.

Si l'accès à ces données représente une opportunité permettant d'améliorer les connaissances, il est important de résoudre des problèmes méthodologiques posés par ces bases de données.

Comment analyser des données « massives » ?

Ce problème s'est posé lors du développement de la génomique en oncologie. Le nombre très important de données et de tests mettait en évidence des mutations génétiques sans lien avec la pathologie concernée.

En effet, le nombre très important de données traitées augmente très fortement la puissance des comparaisons risquant de rendre significatif au sens statistique du terme, des différences faibles. De plus, un nombre important de comparaisons réalisées risque lui aussi de faire apparaître comme significatives des différences en fait aléatoires.

Ainsi, la question n'est pas de mettre en évidence des différences qui apparaîtront très fréquemment mais de valider des hypothèses faites dans la cadre d'un modèle d'effet.

Car les corrections disponibles pour prendre en compte les comparaisons multiples (type Bonferroni) ne sont pas adaptées au nombre très important de comparaison rendu possible par un nombre de données disponibles considérable.

Indicateurs – Open data ou études ad hoc ?

L'ouverture des bases de données de santé pose la question des outils les plus appropriés pour évaluer les politiques publiques.

Les indicateurs, conçus pour monitorer une intervention ou un programme représentent actuellement, le gold standard pour l'évaluateur. La mobilisation d'indicateurs, conçus pour leurs qualités métrologiques (SMART) au regard de la question posée nécessite des moyens et une anticipation souvent incompatibles avec le calendrier de la mise en œuvre des programmes.

Les études ad hoc sont une alternative à des indicateurs recueillis en routine. En effet, si elles ne peuvent être effectuées qu'en un nombre limité d'itération, elles sont conçues et

réalisées pour répondre à une question spécifique et limite de ce fait les biais d'interprétation.

L'analyse des données issues des grandes bases de données est potentiellement la moins contraignante puisqu'elle ne nécessite pas de planification particulière par rapport à la mise en œuvre de la politique à évaluer

Que signifient les variations observées ?

Conçues pour un autre objet que l'évaluation, les grandes bases de données comportent dans leur conception des lacunes pouvant soit biaiser soit interdire leur utilisation pour une évaluation : description grossière des maladies rares par exemple dans les bases du SNIRAM ou du PMSI, identification avec le n° de sécurité sociale de l'ayant droit plutôt que n° INSEE, etc.

La consommation (au sens large) ou le remboursement sont dans la plupart des cas l'objet des grandes bases de données maintenant accessibles. Le lien entre consommation et état de santé est soumis à l'action de nombreux facteurs de confusion.

Il est donc nécessaire de constituer un modèle logique de l'effet attendu avant d'analyser les variations dans ces grandes bases de données pour limiter les biais d'interprétations.

Prospective : quel avenir pour les données de santé 3.0 ?

Si les grandes bases de données publiques recueillent pour l'essentiel des données de maladies et non des données de santé, les données de santé sont de plus en plus stockées par des opérateurs privés.

Les principales données concernant les comportements alimentaires sont ainsi recueillies par les grandes enseignes au moyen de leur carte de fidélité. Demandée volontairement par les consommateurs elles permettent de relier les données de consommations à de nombreuses caractéristiques personnelles et familiales, et ce en échange de très faibles avantages au regard du nombre de données fournies.

C'est le cas aussi des données recueillies par les objets connectés (podomètres sur smartphones, montres connectées recueillant les phases de sommeil ou le rythme cardiaque). Il en est de même des données recueillies lors des consultations des sites internet (de santé entre autres) par les CRM (« Crew Resources Management »). Ces données couplées avec des données de géolocalisation ou des adresses IP ou des numéros MAC, recueillies volontairement par les utilisateurs vont représenter une formidable description de l'état de santé d'une population.

Les Big Data de santé sont devenues un enjeu majeur de développement tant pour l'industrie pharmaceutique que pour les chercheurs. Ainsi début septembre, le laboratoire pharmaceutique français Sanofi a signé un partenariat avec Google pour la surveillance des patients diabétiques grâce à un dispositif connecté. L'institut Gustave Roussy, premier centre de lutte contre le cancer d'Europe, vient de signer un partenariat avec Dell pour renforcer les capacités de traitement des données issues des séquençages. Autant de données de santé qui seront stockées et traitées par des opérateurs privés et potentiellement hors du champ de l'évaluation des politiques publiques de santé.

Quel sera, en effet, le statut de ces données ? Feront-elles l'objet de négoce comme cela a été le cas pour les données de génétique (certains établissements hospitaliers vendant les données génétiques de leurs patients à des laboratoires pharmaceutiques) ? Dans quelle mesure les pouvoirs publics auront-ils accès à ces données ? Alors que des données publiques longtemps protégées deviennent accessibles aux chercheurs et aux évaluateurs, d'autres, tout aussi stratégiques pour l'évaluation des politiques publiques, notamment dans le domaine de la prévention, sont recueillies massivement sans que leur utilisation dans un but de santé publique soit possible à terme, de façon certaine.

1.3. Les données de santé : des exigences.

Christian Babusiaux, président de chambre à la Cour des comptes, président de l'Institut des données de santé.

La mobilisation des données, un enjeu pour l'efficacité des politiques de santé et d'assurance maladie, et un enjeu pour les évaluateurs des politiques publiques, au premier chef la Cour. Les données, une chance pour l'évaluateur mais qui doit le conduire à une exigence accrue vis-à-vis des producteurs de données. La révolution des données de santé, une chance potentielle pour ceux qui suivent et évaluent les politiques de santé.

1. La nécessité particulière des données dans le domaine de la santé : nombre considérable et extrême diversité des acteurs du système de soins, nombre considérable de pathologies et de traitements. En outre, la santé n'est pas seulement liée aux soins : elle est la résultante d'éléments multifactoriels : socio-professionnels, héréditaires, génétiques, alimentaires, comportementaux, environnementaux. Elle ne peut être comprise que par le rapprochement et le croisement de données très diverses.

2. Le cloisonnement des acteurs (multiplicité des agences sanitaires, etc.), celui des systèmes d'information (entre ville et hôpital, et même entre hôpitaux voire entre services) ont longtemps constitué un frein à la connaissance.

3. Les progrès de la période récente : « chaînage » des données entre ville et hôpital, élargissement du SNIIRAM (le Système national d'information inter-régimes de l'Assurance maladie) à tous les régimes, allongement de son historique (8 ans et demi aujourd'hui, 9 ans et demi en 2016), expérience réussie de « chaînage » entre remboursements par les régimes obligatoires et les régimes complémentaires (MONACO) ; la constitution de grandes cohortes sans commune mesure avec celles qui les ont précédées (INCA base cancer) etc.

Une masse de données et de possibilités de traitement en croissance exponentielle.

Aux bases publiques les plus souvent évoquées (PMSI, SNIIRAM) s'en ajoutent d'autres (dossier pharmaceutique, base de données des aliments de l'ANSES...). De grandes bases privées (assureurs complémentaires, plates-formes santé). Des objets connectés et big data : des possibilités nouvelles de connaissance des comportements réels de santé et de rapprochement d'éléments traduisant le caractère multifactoriel de la santé, qui ouvrent la voie à une évaluation moins réductrice.

Les progrès nécessaires dans les bases publiques pour donner des moyens améliorés au suivi et à l'évaluation des politiques

Outre la question de l'ouverture des bases, qui ne concerne pas directement certains évaluateurs et notamment la Cour, qui peut, de droit, accéder aux données des bases publiques si elle le veut, ne pas méconnaître les limites de ce qui existe et pousser aux améliorations nécessaires.

1. Les imperfections du contenu des bases : cf. les écarts considérables dans le nombre de décès d'origine alimentaire entre le PMSI et le registre de décès ; l'absence de bases sur le handicap (malgré la loi de 2004) et sur la dépendance ; le caractère non représentatif de l'échantillon généraliste de bénéficiaires (EGB) et l'absence de couverture de domaines majeurs de l'activité médicale et hospitalière.
2. La facilitation nécessaire de l'appariement des bases par l'utilisation du NIR (tout dépendra du décret d'application de l'article 47 de la loi santé).
3. Des bases qui concernent les remboursements mais pas tous les soins et pas la santé elle-même, contrairement à ce qui existe dans d'autres pays

Des efforts de transformation à conduire par l'évaluateur lui-même.

Les limites actuelles dans l'utilisation des données par les agences publiques.

Dans l'utilisation des bases de données publiques, même quand ces instances y ont accès (cf. pour l'accès aux données exhaustives du SNIIRAM ouvert à l'InVS depuis 2012 et à la HAS depuis 2013). Seule autorité à avoir commencé à les utiliser couramment et à s'être équipée et organisée en conséquence : l'ANSM.

Dans la mobilisation des données de type « big data ».

Dans la quasi-absence d'accès pour les ARS malgré leur rôle dans la mise en œuvre et le suivi des politiques de santé

Globalement, dans ce domaine, dans une sorte de retard technologique d'une grande partie de la sphère publique par rapport à la rapidité de plus en plus forte des progrès de la sphère privée.

La problématique à la Cour.

1. La Cour manie beaucoup de chiffres mais ne s'est pas encore vraiment lancée dans la mobilisation des données.

Elle a commencé récemment à s'organiser, avec le Centre d'appui métiers, mais ce n'est, chacun en est conscient, qu'un premier pas.

2. Comment passer de l'idée à sa mise en œuvre ?

Il y faut de la formation, par exemple pour qu'un nombre suffisant de rapporteurs et vérificateurs sachent les types de données existantes, par exemple à la DREES ou à la

CNAMTS, et comprennent, au moins de manière générale, ce qui peut en être tiré pour les enquêtes.

Mais, au-delà, il faut une véritable acculturation. On ne peut pas penser à une utilisation seulement épisodique et ponctuelle. Le besoin est permanent, structurel, et on ne peut d'ailleurs connaître les bases et les utiliser que par une utilisation régulière. De même, il est sans doute impossible que quelqu'un puisse être compétent pour utiliser des bases dans des domaines très différents et soit mobilisé à l'occasion pour utiliser celles, aujourd'hui complexes, du domaine de la santé. Il faut une « expertise embarquée ». Certes, il ne faut surestimer ni la difficulté d'utilisation du PMSI ou du SNIIRAM, ni le nombre de personnes nécessaires, mais ce sujet rappelle ce qu'a connu la Cour lorsqu'elle a dû mettre en œuvre la certification des comptes : elle a su alors recruter en conséquence ce que nous avons appelé les « experts ».

Cependant, acquérir une expertise sur les données elles-mêmes ne suffit pas car elles ne peuvent être utiles que si on sait leur donner un sens et les interpréter. Il faut donc progresser simultanément dans la capacité d'expertise médicale et sociale ou du moins dans le niveau de connaissance de ce qui est le fond des sujets. C'est donc un progrès qualitatif global qu'il faut rechercher pour mobiliser pleinement les données au service du suivi et de l'évaluation des politiques.

Il faut aussi intégrer la préoccupation de la mobilisation des données le plus en amont possible du processus de travail, c'est-à-dire au moment de la programmation et à celui de la note de faisabilité pour intégrer dès ces phases ce qui sera nécessaire en temps, en délais, en préparatifs de l'enquête pour mobiliser des données. Après, cela risquera d'être trop tard.

Au total, les données sont, dans le domaine de la santé, une chance de donner une profondeur nouvelle au suivi et sa véritable dimension à l'évaluation des politiques. C'est aussi une contrainte parce que, si l'évaluateur ne s'adaptait pas, il risquerait fort de se trouver rapidement dépassé par les progrès des autres acteurs, publics et privés de ce domaine. Mais attention aux illusions. Il ne suffirait pas de penser que les choses viendront d'elles-mêmes. C'est un processus de transformation, à vrai dire passionnant, qui est à réaliser, et cette maison, qui a connu tant d'évolutions déjà, y compris pour la certification, que j'évoquais tout-à-l'heure, trouvera dans cette adaptation la chance d'une évolution nouvelle.

Chapitre 2. Seconde table ronde : les mobilisations des données pour le suivi et l'évaluation des politiques publiques dans le domaine du logement.

Maurice Baslé, professeur émérite des universités, CREM-CNRS-Université de Rennes 1 et Chaire Connaissance et action territoriale à l'Université de Bretagne Sud.



Le domaine du logement est probablement un des domaines les plus observés et qui donne lieu à nombre de réutilisations par les professionnels ou les administrations qui appliquent les prescriptions contenus dans les programmes publics. La nécessité de plateformes de données interopérables en vue de ces réalisations est le défi du moment (voir Chaire ACT-TER).

Le domaine du logement est aussi celui qui fait l'objet d'usage des données par les chercheurs. Des chercheurs aujourd'hui qui changent de paradigme lorsqu'ils ont accès à des masses de données considérables puisque, du traitement du « big data », du rapprochement des sources de données sur des plateformes les rendant interopérables, peuvent jaillir de nouvelles questions de recherche.

Un récent rapport d'inspection a montré la diversité mais aussi le besoin de faire la réorganisation de ces ressources considérables. 34 recommandations qui, si elles étaient suivies, pourraient en effet permettre de mieux mobiliser des données au service de leurs

usagers au premier chef les évaluateurs de politiques publiques mais aussi les chercheurs appliqués.

Voir le rapport d'inspection sur les statistiques du logement établi par Alain Chappert Inspecteur général de l'Insee Jacques Friggit Ingénieur général des ponts, des eaux et des forêts Isabelle Kabla-Langlois Inspectrice générale de l'Insee Patrick Laporte Inspecteur général de l'administration du développement durable - JUIN 2014
<http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000532.pdf>

2.1. La mobilisation des données au SOES (Service de l'observation et des statistiques) du Commissariat général au développement durable.

Guillaume Houriez, sous-directeur des statistiques du logement et de la construction au SOES - ministère du logement.

Les questions à aborder sont les suivantes. Quelles données pour les évaluations ? Les données au SOeS doivent être utiles à l'évaluation. Se pose la question de l'accès aux données produites par le SOeS, celle des évolutions envisagées pour améliorer l'accès aux données et celle des progrès à réaliser dans la conception des systèmes d'information.

Pour être utiles aux évaluateurs, les données doivent remplir plusieurs critères :

- La fiabilité des informations : description de la source, transparence des méthodes.
- La pérennité des informations pour évaluer les impacts dans le temps.
- Des données finement géo-localisées pour évaluer des politiques qui ont souvent une dimension spatiale.

Plusieurs catégories de données sont à disposition.

- Les données de cadrage : le compte du logement.
 - Séries macroéconomiques depuis 1984.
 - Place du logement dans l'économie du logement, détail des aides au logement.
- Les bases administratives exhaustives pérennes et géo-localisées.
 - RPLS : description du parc locatif social (pas d'information sur les occupants).
 - Filocom : fichier d'origine fiscal qui comporte des informations sur le logement et sur leurs occupants, notamment leur revenu imposable (demain : RSL...).
- Les enquêtes statistiques exhaustives pérennes et en cours de géolocalisation.
 - Sitadel : suivi de la construction neuve (autorisation de construire et mises en chantier), enquête mensuelle.
 - ECLN : Commercialisation des logements neufs par les promoteurs immobiliers (volumes, prix) ; enquête trimestrielle.

- EPTB : Enquête Prix des maisons individuelles et des terrains à bâtir, enquête annuelle.
- Des enquêtes thématiques.

Phébus : enquête sur les consommations d'énergies des ménages et sur la performance thermique des logements.

Nota bene. Le SOeS n'est pas le seul producteur de données sur le logement. Voici quelques exemples d'autres sources. Les bases notariales : informations sur les prix des transactions, le logement, les acheteurs et les vendeurs. Elles sont achetées par la DHUP et le SOeS tous les deux ans. Les sources produites par l'Insee : recensement de la population, enquête nationale logement, enquête loyer et charges, etc... D'autres sources de données : celles des observatoires locaux des loyers (la diffusion via le CASD pour les chercheurs est prévue). Les données administratives du ministère du logement : financement du logement social (Sisal), la demande de logement social (système national d'enregistrement ou SNE), <http://sne.info.application.territoires.gouv.fr/> etc.

L'accès aux données.

Le secret statistique (loi 51-711 du 7 juin 1951). Il s'impose à toutes les enquêtes statistiques. Ne permet pas de diffuser des données non anonymisées.

Le cas particulier de Filocom. La convention qui nous lie avec la DGFIP limite les accès, même aux agrégats secrétisés.

Des agrégats détaillés (à la commune) pas systématiquement diffusés. Sitadel est diffusée à la commune, certaines variables du RPLS le sont également ; mais les autres sources sont diffusées à la région ou au département.

Le Soes accueille des chercheurs dans ses locaux et ils travaillent dans le cadre de conventions : écoles d'économie de Paris et de Toulouse notamment.

La diffusion des données individuelles de RPLS, encadrée par un décret spécifique, est-elle assez large ? En fait, toutes les administrations ont accès aux données individuelles.

Les évolutions en cours.

Diffusion plus systématique des données d'enquête à la commune. Cela concerne le répertoire du parc locatif social, l'enquête sur la commercialisation des logements neufs et l'enquête sur le prix des terrains à bâtir.

Diffusion sur le réseau Quetelet et sur le CASD. Enquête EPTB notamment.

A noter que le fichier de diffusion du répertoire statistique du logement (RSL, Insee) qui devrait succéder à Filocom sera disponible via le CASD.

Dans le futur, il est possible d'envisager de relâcher le secret statistique lorsque les parties prenantes sont d'accord. Le SOeS s'engagera dans cette démarche en 2016 dans le cadre de la refonte de l'enquête sur la commercialisation des logements neufs en collaboration avec les promoteurs immobiliers et le Comité du secret.

Quelques marges de progrès subsistent. Les sources qui ont été présentées ne sont pas conçues aujourd'hui spécifiquement pour des besoins d'évaluation. Ce sont des sources généralistes qui

répondent à des besoins multiples. Il faudrait mieux penser aux systèmes d'information lorsqu'on conçoit les politiques publiques. Il ne s'agit pas de développer un système d'information ou une enquête statistique ad hoc. Mais il faut davantage penser les outils de gestion comme des systèmes d'information. Et améliorer l'interopérabilité des systèmes. Un exemple : RPLS et Sisal pour pouvoir suivre le logement social de son financement à sa mise en service. Mise en œuvre prévue mi-2016.

2.2. Les données mobilisables par l'auditeur public : l'exemple du logement social.

Christian Galliard de Lavernée, conseiller maître en service extraordinaire à la Cour des comptes.

Le projet d'évaluation de la politique publique du logement social par la Cour des comptes a fait l'objet d'une étude de cadrage au premier semestre 2015. Elle s'est penchée notamment sur l'accès aux bases de données du logement pour asseoir les analyses quantitatives nécessaires à cette enquête.

Depuis son approbation début juin, des contacts ont été pris par les rapporteurs avec les principaux producteurs de données, avec l'appui du Centre d'appui métier de la Cour, qui développe sa compétence dans le domaine de la collecte et du traitement des données.

La validation des questions évaluatives détaillées par la formation des juridictions financières en charge de cette évaluation, installée le 9 octobre, permettra de faire passer en phase opérationnelle le dialogue avec les fournisseurs de données, principalement l'INSEE, la Caisse nationale d'allocations familiales et le Service d'observation et d'études statistiques (SOeS) du ministère de l'écologie.

L'enjeu est de réussir à catégoriser l'occupation et la demande de logement social en fonction du revenu fiscal ou du revenu disponible des ménages, d'apparier ces données avec la taille et le loyer des logements, et de proposer tous ces chiffres aux parties prenantes à l'échelle de territoires tests définis par une unité-urbaine, voire à l'échelle du quartier ou de l'IRIS quand on analysera la ségrégation spatiale des revenus comme indice de la mixité sociale.

Quels sont les premiers enseignements de ce travail à peine engagé ?

- 1) Sur la protection des données
 - Les questions du secret statistique et de la protection des données individuelles peuvent freiner l'ouverture des données, même pour la Cour des comptes.

L'article L.141 du code des juridictions financières, sur lequel la Cour s'appuie pour obtenir communication des données nécessaires à ses enquêtes, ne dit pas que les secrets protégés par la loi, par exemple le secret statistique (loi de 1951 et règlement européen de 2009) ne peuvent être opposés à la juridiction. Son droit de communication doit s'exercer dans les conditions fixées par les textes et en accord avec les autorités de protection (comme c'est le cas du secret de l'instruction ou du secret fiscal).

La loi Informatique et liberté de 1978 conduit au même type de solution : la Cour a le statut de tiers autorisé, mais ne peut exiger des données nominatives détenues par des

organismes non soumis à son contrôle que de façon motivée et sur des personnes nommément désignées. L'exploitation de fichiers nominatifs nécessite donc des procédures de protection convenues avec le détenteur (anonymisation), et ce n'est que très récemment que la Cour peut demander communication de données individuelles relatives à des catégories de personnes (art. L81 de la loi de finances rectificative du 29 décembre 2014, ouvrant ce droit aux services fiscaux, et bénéficiant à la Cour sur le fondement de l'art. L.141-5 du CJF).

- Les règles de protection appliquées par les organismes en réponse à la Cour ne sont pas exactement les mêmes, en l'absence de référence législative ou réglementaire vérifiée.

L'anonymisation s'impose, dès lors que l'on ne fait pas partie du service public de la statistique. Mais la « secrétisation » des données, qui consiste à occulter dans les tableaux statistiques les cases qui concerneraient un faible nombre de sujets d'enquête, est un domaine flou : elle concerne une population de moins de 11 personnes au SOeS, et 5 personnes à l'INSEE.

Le partage d'une doctrine commune sur le secret partagé entre le producteur des données et l'auditeur public n'est pas abouti.

2) Sur les moyens de traitement

Les données ouvertes invitent à des investigations inconnues jusque-là. Indépendamment du travail des auditeurs, ce progrès a un coût : celui des capacités de traitement ou de transmission mobilisées, et celui des heures de travail des statisticiens ou informaticiens qui s'y consacrent.

Là aussi, les règles de prise en charge de ces frais supplémentaires ne sont pas bien établies. On rencontre des questions de facturation des prestations, de mise à disposition de personnel qualifié par le commanditaire, ou prosaïquement de délais de réponse menant à l'impasse faute de moyens chez le producteur.

La Cour développe un pôle Données au sein du CAM, et des conventions avec les fournisseurs de données, de façon à résoudre de concert les difficultés objectives rencontrées et de normaliser progressivement les solutions adoptées.

3) Sur la relation entre les producteurs et les utilisateurs

L'ouverture des données modifie en profondeur la définition des rôles : dans une certaine mesure, la production de statistiques échappe au monopole de ceux dont c'est le métier. Cela peut avoir un impact subjectif sur leur identité professionnelle, nourri par un risque objectif de perte de la qualité et de la sécurité des données diffusées. Pour autant l'intervention croissante des utilisateurs dans les choix de production de données est évidemment féconde, à condition qu'ils formulent de façon claire et sélective leurs spécifications.

Ce croisement des légitimités appelle des échanges renforcés entre les deux métiers dont la définition est moins étanche. La discussion des conventions de travail développées par la Cour et déjà évoquées a pour effet de renforcer le respect mutuel entre partenaires, et

d'identifier par l'apprentissage réciproque les contraintes nées du développement de la société de l'information.

Bien sûr, le recours aux données chiffrées n'est pas la panacée de l'évaluation. L'approche qualitative garde sa nécessité et même sa primauté : au-delà des réalités quantifiables, il faut décrire le ressenti des acteurs, leurs comportements et leurs jeux de rôle.

Mais le partage des données au cours du processus d'évaluation est certainement un facteur d'ajustement spontané des jugements et des comportements. Il favorise la convergence des points de vue et la coopération des acteurs, car la preuve met fin aux faux débats. De ce point de vue, la juste description modifie les situations car elle fait naître l'action. C'est ce que les mois qui viennent confirmeront sans doute à l'équipe en charge de l'évaluation de la politique du logement social.

Henri Verdier. DINSIC.

On le voit, en effet, le champ des possibles est infini. Une question intéressante à cet égard consistera à nous demander si nous nous sommes bien préparés à saisir cette révolution.

On peut noter en particulier la nécessité, de plus en plus impérieuse, de concevoir, en même temps qu'une politique publique, le dispositif de données qui permettra de l'évaluer et donc de la faire entrer dans un processus d'amélioration continue, et de la piloter avec finesse. A cet égard, de même qu'il n'est plus acceptable de transférer des compétences sans transférer en même temps les ressources nécessaires à l'exercice de ces compétences, il conviendra désormais de se demander, avant tout transfert de compétence, si les données nécessaires à l'exercice de cette compétence ont bien également été transférées.

Au-delà de cette question organisationnelle, il conviendra d'apporter la plus grande attention à la question centrale de la gouvernance de la donnée, qui fait l'objet du rapport de l'Administrateur général des données qui sera publié début 2016.

Nos services publics ne se sont en effet pas encore organisés pour recenser les trésors de données dont ils disposent, ni pour en organiser la meilleure circulation possible – dans le respect des secrets légaux – en leur sein, voire à destination de la société civile lorsque cela est possible (open data).

Conclusion.**Pierre Savignat, président de la SFE, société française pour l'évaluation.**

Nous devons nous orienter vers un processus intégré des données au sein des politiques publiques, depuis la conception jusqu'à l'évaluation en passant par la mise en œuvre.

La production et le traitement de données fiables et pertinentes sont nécessaires à toute évaluation. De ce point de vue, la qualité de l'appareillage de suivi, de collecte et de traitement des données au fil de l'eau facilite grandement le processus évaluatif.

Par ailleurs, l'évaluation concourt à mettre ces données en perspective, à enrichir leur utilisation, à rendre compte de la complexité de leurs significations, à approcher les relations causales. Autrement dit, le processus évaluatif éclaire les dimensions de sens des données utilisées au vu des objectifs de l'action publique. Le croisement des données avec des approches qualitatives, avec le recueil des points de vue des parties prenantes permet de consolider l'analyse, d'étayer l'appréciation et d'éclairer solidement la décision publique.

Enfin, l'évaluation peut contribuer à apprécier les systèmes de données et à les faire évoluer.