La protection des données personnelles en assurance dialogue du juriste avec l'actuaire

Delphine Cocteau-Senn[1], Arthur Charpentier[2], and Rodolphe Bigot[1]

[1]Université de Picardie Jules Verne (CEPRISCA, EA3911)
[2]Université de Rennes (CREM, UMR6211)
Affiliation

Les données, instruments de mesure du risque

De la donnée à la donnée « à caractère personnel »

DCS[1] : Depuis toujours, les données de l'assuré sont au cœur de la relation instaurée par le contrat d'assurance. Elles servent la mesure du risque individuel, objet du contrat et déterminent tant la décision de l'assureur de prendre ce risque en charge que sa tarification. De son côté, l'assureur se trouve dans un cas d'inversion du cycle de production : il doit évaluer le coût réel de son produit par le biais d'analyses prédictives du coût du risque et fait appel à l'actuariat. Ces analyses reposent sur des observations statistiques, et donc d'un ensemble de données massées que l'on nommera « données actuarielles ». Dans cette perspective, il apparait que plus l'information est riche, plus les critères sont fins, mieux l'assureur pourra affiner la mesure des risques que sa mutualité prendra en charge et meilleure sera la coïncidence entre son offre de couverture et les besoins individuels de l'assuré (segmentation de l'offre). Or, dans une perspective concurrentielle, la segmentation est devenue indispensable pour pallier les effets néfastes du phénomène d'antisélection[2]. Cette circonstance ne peut que pousser l'assureur à chercher toujours plus de données.

Depuis le renforcement des préoccupations sur la protection des données, et notamment en raison de l'entrée en vigueur imminente du Règlement pour la Protection des Données Personnelles (dit RGPD) [3], l'attention se focalise sur le caractère « personnel » de la donnée. La définition qu'en donne le Règlement est large, suivant sur ce point un chemin emprunté il y a déjà longtemps par le législateur français[4], mais ne coïncide pas toujours avec la compréhension plus étroite qu'en ont les acteurs de terrain ou les intéressés eux-mêmes, pour la majorité desquels la donnée personnelle se cantonnerait à la vie privée ou intime.

Au sens du nouveau texte européen, l'expression « donnée à caractère personnel » vise « *toute information se rapportant à une personne physique identifiée ou identifiable* (ci-après dénommée «personne concernée») ; est réputée être une «personne physique identifiable» une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale » (RGPD, art. 4.1). La protection assurée par les textes va ainsi bien au-delà de la simple donnée « nominative » mais vise tout élément permettant l'identification, même indirecte[5].

Question à l'actuaire :

Sur quel type de données l'actuaire d'assurance travaille-t-il dans sa pratique de modélisation des risques, et cela comprend-il des données directement ou indirectement identifiantes, soit des données dites « à caractère personnel » ? Peut-être même des informations sensibles, comme la race ou la religion ? Sous quelles formes vous parviennent ces données ? Et surtout, pour éclairer l'amalgame souvent fait entre la notion de donnée à caractère personnel et celle de vie privée, quelles en sont l'objet et les sources ?

---

[1]Ce travail prend pour point de départ les échanges ayant eu lieu avec Arthur Charpentier lors de la table ronde relative aux données d'assurance lors du colloque « Droit des données personnelles » (Amiens, 7-8 nov. 2016) qu'il a pour but de développer tout en conservant la forme originaire du dialogue. Le débat s'est également enrichi de la contribution de notre collègue R. Bigot, que nous remercions vivement d'avoir accepté de se joindre à la discussion.

[2]En assurance, la théorie économique montre que si les agents sont rationnels et si l'assurance n'est pas obligatoire, les « mauvais risques » ont un intérêt supérieur à la moyenne à souscrire un contrat d'assurance (phénomène dit de l'antisélection).

[3]Règlement (UE) 2016/679 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel, dont l'entrée en vigueur est prévue pour le 18 mai 2018.

[4]V. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, dite « Loi Informatique et Libertés ».

[5]C'est ainsi que la CJUE a pu considérer dans un arrêt du 19 oct. 2016 (affaire C-582/14) que l'adresse IP dynamique d'un internaute conservée par le FAI constituait une donnée personnelle au sens de la directive 95/46/CE du 24 octobre 1995, dès lors qu'elle elle permet de l'identifier une fois associée à une autre information (adresse mail, horaires de connexion, etc.).

L'assureur, au premier chef imagine-t-on s'agissant des données fournies par l'assuré, les données publiques, les réseaux sociaux ?

AC : La source première de données dont dispose l'actuaire pour modéliser les risques, et construire un tarif segmenté est la base constituée à partir des informations collectées dans les formulaires de souscription. Ces formulaires contiennent un numéro de police, le nom de l'assurée, son adresse, sa date de naissance, etc. Les actuaires ont souvent accès aux données brutes, directement. Le nom et le prénom peuvent être exclus, mais en assurance habitation, il est souvent utile d'avoir accès au lieu précis d'habitation : l'adresse, et l'étage (pour les immeubles). Ces données sont bien souvent utiles car le risque dépend du quartier, mais aussi de l'étage : pour le cambriolage savoir si l'habitation est au rez-de-chaussée est important, pour le dégât des eaux, c'est souvent savoir si l'habitation est au dernier étage. En assurance automobile, on peut utiliser des informations relatives au lieu d'habitation (habiter en banlieue ou en campagne n'impose pas le même genre de conduite) donc le code postal est souvent utilisé, mais aussi le modèle de véhicule (indiqué sur la carte grise), et l'âge du conducteur principal. En regroupant ces trois variables dans une commune de quelques milliers d'habitants, la personne est bien souvent identifiable.

A partir de ces données provenant des questionnaires, il n'est pas rare de procéder ensuite à des croisements de données. Par exemple à partir du modèle du véhicule, on peut trouver sa cote à l'argus (ce qui donne un montant maximal de remboursement en cas de dommage matériel), sa puissance (dans certains pays, seule la puissance du véhicule est utilisée comme variable tarifaire), le nombre de places, la marque (certaines marques ont des coûts de réparation plus élevés), etc. En assurance habitation, l'adresse permet d'avoir toutes sortes d'informations. Croisées avec des données du cadastre, on peut avoir l'âge du bâtiment, mais on peut aussi voir des informations sur le quartier (nombre de cambriolages par exemple). Depuis quelques années, les assureurs réfléchissent à l'utilisation de données « connectées », comme les bracelets qui mesurent le rythme cardiaque ou le nombre de pas faits dans une journée, et les boitiers GPS dans les véhicules. Ces données sont intéressantes pour comprendre le risque, mais plus difficilement à des fins tarifaires. En effet, la prime d'assurance est fixée ex-ante, et ces données sont collectées ex-post. Une solution peut-être de faire une offre commerciale indexée sur un engagement de l'assuré, vérifiable par ces données connectées : offrir un rabais de 15% si la personne s'engage à faire en moyenne sur une semaine au moins 10 000 pas par jour, ou moins de 7000 km avec le véhicule sur une année. Les boitiers GPS permettent en théorie d'avoir accès à énormément d'information sur l'assurée (lieu du domicile, localisation du stationnement, lieu du travail, lieu de l'école des enfants et du club de sport, etc.). Mais les assureurs sont tributaires de données externes, fournies par le fournisseur du boitier GPS. Pour des assurances de véhicules commerciaux, certains assureurs demandent expressément à ne pas avoir d'information sur la vitesse des véhicules par exemple. Certains autres demandent à n'avoir accès qu'à des informations très synthétiques sur la conduite : nombre de trajets, nombre de kilomètres, temps de conduite la nuit, etc. Ces données sont alors non-identifiantes, contrairement à nombre de données de télématiques[6].

Enfin, pour les données sur les réseaux sociaux, c'est plus sensible. Plusieurs études (aux États-Unis et en Angleterre) ont montré que l'utilisation d'informations relatives aux réseaux d'amis était très prédictive d'un défaut ou d'un retard de remboursement de crédit hypothécaire. On peut imaginer aller encore plus loin en regardant le contenu de ce qui est mis en ligne. Là aussi des études ont montré que les photos publiées sur une page Facebook pouvaient être utilisées pour prévenir le suicide. Lire le contenu permet probablement d'avoir des informations sensibles. En lisant les tweets précédant une élection présidentielle, il est possible d'avoir une prévision (avec une probabilité assez élevée) des orientations politiques de l'assuré. Mais rares sont les études qui montrent un lien entre les orientations politiques, sexuelles, religieuses d'une personne et son nombre de dégâts de eaux, ou d'accidents de la route, donc rares sont les actuaires à regarder ce genre de variables. En revanche, regarder les réseaux d'amis sur Facebook est souvent utilisé lors d'études sur la fraude.

Quand un actuaire fait un tarif, c'est un exercice de statistiques prédictives : en utilisant les informations passées, on essaye de voir si les assurés qui avaient des caractéristiques proches ont eu ou pas des accidents (dans le passé), combien, à quel coût. Compte tenu des délais de déclaration et de clôture des sinistres, il faut souvent un long historique. Par exemple pour estimer le coût potentiel d'un accident corporel, pour les contrats d'assurance automobile, on ne peut pas se limiter aux statistiques relatives aux cinq dernières années : les plus gros sinistres sont encore ouverts, certains patients (les états les plus graves et donc les plus coûteux) sont encore dans un état non-stabilisé. Il est alors indispensable d'utiliser les données les plus anciennes possibles, avec la difficulté de tenir compte d'amélioration techniques sur les véhicules améliorant la sécurité, les changements de conduite (radars automatiques incitant à réduire globalement la vitesse) et l'inflation (hospitalière et juridique) pour les sinistres relativement anciens. Les actuaires « recyclent » en

[6]Sur les données géolocalisées de téléphones cellulaires, certaines études ont montré que 4 points (lieux et heures approximatifs) suffisent à identifier 95% des individus dans une base de données (de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. Nature srep. 3, 1376; DOI:10.1038/srep01376 (2013).

permanence les anciennes bases de données.

De la donnée personnelle à la « donnée sensible »

DCS : Le terme de « données sensibles » n'est pas anodin pour le juriste. Un régime particulier est en effet réservé à ces données, dont le traitement n'est autorisé que très exceptionnellement (article 9). Mais qu'entend-ton exactement par « données sensibles » ? Sont visées par-là les données dont le traitement « révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique ». Il faut donc considérer qu'il y a des données directement sensibles du fait de leur contenu, à côté de données qualifiées « sensibles » *indépendamment de l'objet direct de l'information qu'elles portent*[7] du simple fait qu'elles permettront de déduire l'information jugée sensible. L'assureur peut être destinataire de données dont l'information est directement sensible (ex. assureur santé ou invalidité), mais il l'est assurément de toute une série d'autres, non sensibles en elles-mêmes, qui sont cependant susceptibles d'être croisées entre elles et de révéler alors bien plus. Si l'on songe à la masse des informations susceptibles de figurer dans un dossier d'assuré, sans compter celles qui peuvent être déduites des données très fines des objets connectés, la marge de manœuvre de l'assureur semble étroite.

Questions à l'actuaire :

Dans les données d'assurés sur lesquelles travaillent actuellement les actuaires, y a-t-il des données « sensibles » au sens du RGPD ? Ces résultats ou celui du croisement des données vous semblent-ils suffisamment fiables ou éloquents pour que l'on puisse considérer que de nombreuses données apparemment anodines pourraient tomber dans le champ de la donnée « sensible » du fait de ce qu'elles sont « susceptibles de révéler » (ex : opinion politique ou religieuse) ?

AC : Comme évoqué auparavant, un actuaire ne cherchera pas ces données, car aucun modèle n'a établi de relation causale entre les opinions politiques, les croyances religieuses, etc., et la sinistralité. Mais le fait est qu'il est aujourd'hui possible d'avoir accès à des données très informatives sur ces variables dites « sensibles ». Une question est de savoir si ce phénomène est nouveau. Avec l'adresse précise, on pouvait consulter les résultats électoraux par bureau de vote (voire par urne), et affirmer qu'une personne avait 65% de chances d'avoir voté pour tel ou tel candidat. Avec le prénom de la personne, je peux aussi prédire qu'un certain « Jean-Pierre » a 87% d'avoir plus de 50 ans (à partir de statistiques sur le prénom). Je peux affiner ma probabilité en croisant avec son type de véhicule (certains véhicules sont possédés par des personnes de tel ou tel âge, en majorité). Avec les données GPS, je vois qu'une personne stationne presque tous les vendredis matins à proximité d'une mosquée. S'il existe des enquêtes sur les pratiques des musulmans, je pourrais affirmer qu'il y a 98% de chances qu'elle soit musulmane. Mais je me trompe peut-être, et cette personne va en fait au club de gym en face de la mosquée, et en plus elle est assidue. Que signifie avoir accès à des « données sensibles » ? Faut-il être certain ?

La certitude est un concept inconnu aux statisticiens, donc à partir de quel seuil peut-on affirmer que l'on a à disposition des « données sensibles » ? On peut aussi penser à cet exemple fameux d'une société qui avait pu affirmer qu'une personne était enceinte à partir de son changement de consommation (observé à l'aide de sa carte d'achat d'une chaîne de grands magasins) alors que cette personne l'ignorait. Quid du fait qu'il est (en théorie) possible d'avoir à des données encore plus précises (et justes) que celles que les actuaires rêvaient d'avoir ? Plusieurs études ont montré qu'il existait une relation très forte et très prédictive entre les infractions et les accidents de la route. Au Canada, le nombre de points sur le permis est une variable très importante pour prédire le nombre de sinistres l'an prochain (avec des effets complexes, en particulier quand une personne a perdu beaucoup de points, elle peut être beaucoup plus prudente qu'une autre ayant les mêmes caractéristiques mais tous ses points, car elle ne souhaite pas perdre son permis). Tout statisticien rêve de croiser ces données, afin de mieux comprendre l'accidentologie de ses assurés. Le danger est que le fichier des points ou des infractions ne contient que des informations sur ce qui a été sanctionné. Or l'intuition dit qu'on préfère assurer la personne qui n'a pas eu de chance et qui s'est fait flasher trois fois 5km/h au-delà de la limite qu'une personne qui refuse les priorités sans se faire prendre, et qui a la présence d'esprit de ralentir avant tous les radars. On pourrait imaginer un score construit à partir des données des boitiers GPS, sur le respect des stops, le respect des limitations de vitesse. En un sens, l'actuaire aurait à sa disposition non pas un fichier sensible (des infractions) mais des données beaucoup plus riches, et probablement plus pertinentes.

La collecte des données personnes de l'assuré

DCS : Traditionnellement, l'assureur recueille les informations relatives à son futur assuré à l'aide du questionnaire rempli à la souscription, autrement désigné comme

[7] Ce que semble bien confirmer la rédaction de l'article 9, lequel vise, à côté de données identifiées comme sensibles par leur objet (données génétiques par exemple), des données qui sont sensibles dans la mesure où elles révèlent des informations jugées sensibles (race, opinion politique, etc.).

« la proposition d'assurance », et le cas échéant, à en collecter d'autres à l'occasion d'une déclaration de sinistre. Mais des pratiques nouvelles, comme le couplage du contrat d'assurance avec un objet connecté, émergent, qui invitent à se pencher sur une collecte de données qui seraient effectuée par ce biais.

Données recueillies via la proposition d'assurance

Des données pertinentes au regard du risque à évaluer. Le RGPD vient de modifier assez profondément les obligations des responsables de traitement et s'affiche comme un instrument venant renforcer les droits des individus[8]. A l'instar de la loi Informatique et Libertés de 1978 et de la Directive de 1995[9], le Règlement définit les fondements sur lesquels peut reposer un traitement de données, dont la collecte est le premier stade. Celle-ci doit être au premier chef consentie, et ce pour une finalité déterminée (art. 6.1, a). Pour traiter des données sans le consentement de l'intéressé, il faut pouvoir se prévaloir d'un autre fondement, ce qui sera notamment le cas si la donnée est « nécessaire [...] à l'exécution de mesures précontractuelles » (RGPD, art. 1.b). De ce point de vue, la collecte des données personnelles de l'assuré par l'assureur semble a priori fondée au titre de la mesure précontractuelle qu'est l'évaluation du risque, et ce donc, indépendamment du consentement de l'assuré.

La collecte d'informations, préalable au contrat d'assurance, doit cependant être envisagée d'un autre point de vue du fait que le recueil d'informations concernant l'assuré relève déjà d'une réglementation spécifique du Code des assurances. En effet, celui qui demande la prise en charge d'un risque est légalement tenu aux termes de l'article L. 113-2, 2 dudit Code de « répondre exactement aux questions posées par l'assureur, notamment dans le formulaire de déclaration du risque par lequel l'assureur l'interroge lors de la conclusion du contrat, sur les circonstances qui sont de nature à faire apprécier par l'assureur les risques qu'il prend en charge ». L'assuré doit donc fournir les informations qui lui sont demandées, la question ne se posant pas vraiment en termes de consentement au traitement de ses données mais plutôt de consentement à la relation d'assurance, laquelle implique ipso facto une obligation de renseigner.

En résumé, si le recueil des données par l'assureur dans la proposition d'assurance apparaît fondé au sens du Règlement en ce qu'elle est nécessaire à une mesure précontractuelle déterminant la relation d'assurance, il est avant tout, au regard du Code des assurances, un droit de l'assureur opposable à l'assuré. Cette dualité de régime soulève notamment la question des limites de la collecte des données personnelles. S'agissant du texte européen, cet aspect relève de l'article 5 qui pose les principes de limitation au regard des finalités et de minimisation des données. Ainsi, et quel que soit son fondement, la collecte – ici des données de l'assuré - n'est licite que dans la mesure où elle sert des « finalités déterminées, explicites et légitimes » (art. 5.1.b) et qu'il s'agit de données « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités » (art. 5.1.c). Le Code des assurances considère, quant à lui, que l'assuré n'a d'obligation de répondre aux questions de l'assureur que dans la mesure où il s'agit de « circonstances qui sont de nature à faire apprécier par l'assureur les risques qu'il prend en charge » (C.ass., art. L. 113-2,2). Dans les deux cas, c'est l'évaluation du risque qui dessine les contours de ce qu'il est possible de collecter via la proposition d'assurance car elle en constitue la finalité. Partant, les données qui ne seraient pas strictement nécessaires à cette évaluation du risque, comme par exemple des données qui ne seraient utiles à l'assureur qu'afin d'améliorer sa relation client[10], ne devraient pas pouvoir être recueillies via le questionnaire qui s'impose au candidat à l'assurance. Mais l'assureur étant celui qui détermine les risques qu'il accepte de prendre en charge (ainsi que les différentes catégories tarifaires au sein des populations concernées, selon des critères choisis par lui), il semble bien être seul maître de la « pertinence » des données qu'il peut solliciter au titre de l'obligation de déclaration, ou collecter de manière licite, sans avoir à solliciter le consentement de la personne intéressée au sens de l'article 5.1 du RGPD. Cela conduit naturellement le juriste à s'intéresser de plus près à la manière dont est effectuée la sélection des informations « pertinentes » pour l'assureur, s'agissant d'évaluer le risque, ainsi que leur traduction dans le formulaire de la proposition d'assurance.

Question à l'actuaire :

Pouvez-vous nous éclairer sur le travail de l'actuaire d'assurance, et notamment comment celui-ci détermine le type de données qui seront érigées en variables per-

[8] Le juriste français est néanmoins enclin à la réserve à cet égard dès lors que la loi n° 78-17 du 6 janvier 1978 a déjà depuis longtemps posé les fondements de la protection actuelle. Par ailleurs, les nombreuses imprécisions qui affectent les définitions du RGPD, ou la multiplication des exceptions aux règles prohibitives, sont autant de failles dans la protection. Ce, d'autant plus que le nouveau principe d'accountability (passage d'un système de contrôle a priori de la CNIL, par le biais des déclarations et autorisations, à un contrôle a posteriori) déplace en pratique l'interprétation de ces notions sur le responsable du traitement, qui n'aura pas toujours les ressources juridiques nécessaires, et pourrait, en tout état de cause, être tenté d'entendre de manière extensive les exceptions qui lui sont favorables.

[9] Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (JO L 281, p. 31).

[10] Cf. remarque de B. Beaume, Chief data scientist (Covea), intervenant lors de la table ronde du 8 novembre 2016, sur le fait que l'assureur sollicite nécessairement ce type de données à coté de celles nécessaires à l'appréciation du risque

tinentes d'un risque, et qui donc permettront à l'assureur de définir les conditions de sa prise en charge (exclusion, couverture sous conditions, tarification) ? Cette « pertinence » de la donnée ne dépend-t-elle pas du fait que vous en ayez disposé avant afin d'étudier son influence éventuelle sur le risque assuré ?

AC : Il y a deux philosophies en modélisation : soit on part d'un modèle structurel, et on estime les coefficients du modèle, statistiquement, soit on met toutes les variables à notre disposition, et on regarde ce qui a du sens. Ce sont les débats qui existent aujourd'hui sur le « big data » et la fin des modèles[11]. Des études épidémiologiques peuvent donner le délai d'incubation de maladies, et donc indiquer les durées d'attente nécessaire pour l'assurance des centres de transfusion par exemple. Des études sur les transports peuvent donner des liens entre la gravité des accidents de la route et l'heure de l'accident. Savoir qu'une personne conduit beaucoup la nuit sera alors une information que l'on cherche à avoir. Des enquêtes de psychologie peuvent établir des liens entre la couleur de la voiture, et le type de conduite. Savoir qu'une personne conduit une voiture rouge pourra alors être informatif. Les études d'ingénierie peuvent établir un lien entre l'orientation d'un toit et sa probabilité d'être détruit lors du passage d'une tempête hivernale (souvent d'ouest en est). Connaître l'orientation de la maison devrait être intéressant pour tarifer une garantie tempête pour un contrat multirisque habitation. On peut alors avoir un modèle structurel en disant que la probabilité d'avoir un sinistre, ou le coût d'un sinistre, doit être fonction de telle ou telle variable. Les approches plus récentes, dites « data-driven » disent qu'il n'est pas nécessaire d'avoir un modèle formel présupposé, et qu'un algorithme de recherche de corrélations suffira. Ces débats se retrouvent actuellement dans tous les domaines de l'intelligence artificielle. Historiquement, pour faire de la traduction, on supposait qu'il fallait connaitre la grammaire, la structure de la langue. Mais les traducteurs automatiques actuels se contentent de chercher dans des corpus énormes des phrases proches, puis de mélanger de manière logique. C'est la même chose pour les jeux d'échec ou de go, ou la conduite automatique : on ne fait pas un modèle formel de conduite ou de mouvements de pièces sur l'échiquier (si le fou menace ma reine, je cherche à la protéger), mais on demande de regarder dans des millions de parties jouées s'il existe des situations semblables, et de regarder le mouvement qui a permis de gagner dans le plus grand nombre de cas.

Aussi, techniquement, on cherche des variables corrélées avec la sinistralité. Mais cela ne veut pas dire qu'on a une relation causale. La recherche de relations causales est fondamentale en assurance, car elle peut permettre de faire de la prévention. Si on sait que ne pas mettre deux verrous sur une porte d'entrée augmente la probabilité d'être cambriolé de 20%, un assureur peut inciter un assuré à poser un deuxième verrou (si le coût est inférieur à la surprime sur quelques années), et donc changer le risque. Mais bien souvent, les actuaires n'établissent pas de relations causales, ils observent juste des corrélations, et racontent alors une histoire causale. Le fait de conduire une voiture rouge n'est pas neutre sur la survenance de sinistres. La couleur est alors corrélée avec la survenance de sinistres. Établir une relation de causalité sera compliqué (et techniquement, la personne au volant ne voit pas la couleur extérieure de son véhicule, et si on repeint à son insu sa voiture, pourquoi cela impacterait sa conduite ?). C'est juste qu'on se raconte que les personnes qui achètent une voiture rouge, et pas grise, ont probablement une conduite particulière. La couleur est alors un « proxy » d'une variable non-observée, qui caractériserait le type de conduite. Mais ce n'est pas la couleur en tant que telle qui nous intéresse.

Le souci est que pour tester si la couleur est liée à la sinistralité, il faut l'avoir dans les données : on ne sait si une variable peut être utilisée dans un tarif que si on l'a collectée. On se retrouve dans un problème d'œuf et de poule : pour savoir si une variable peut être utilisée, il faut l'avoir récoltée au préalable, alors que la récolte est conditionnée par le fait qu'il faut avoir établi qu'elle était utile.

Des données « juridiquement disponibles »

RB : Depuis peu de temps, le législateur a instauré, en assurance emprunteur, un droit à l'oubli pour certains assurés atteints par le passé de graves pathologies[12]. En interdisant ainsi la prise en compte des antécédents médicaux d'une particulière gravité, classés dans une grille de référence, les assurés disposent d'une forme de droit de taire certaines in-

[11] C'est l'idée énoncée il y a 10 ans : C. Anderson "End of Theory: The Data Deluge Makes the Scientific Method" (Wired, juin 2008).

[12] a convention AERAS ou " s'Assurer et Emprunter avec un Risque Aggravé de Santé " est le fruit d'un accord entre les pouvoirs publics, les fédérations professionnelles de l'assurance, la banque et les associations de malades. Son but consiste à améliorer l'accès au crédit des malades. La convention et ses principales dispositions sont transposées aux articles L. 1141-2 à L. 1144-4 du Code de la santé publique. Après un premier avenant signé le 1er février 2011 et entré en vigueur le 1er mars 2011, un second avenant a été conclu le 2 septembre 2015. Il confère un droit à l'oubli aux personnes ayant eu une pathologie cancéreuse, consacré par la loi du 26 janvier 2016 et étendu à des pathologies non cancéreuses (CSP, art. L. 1141-5). Lors de la souscription d'un contrat assurance emprunteur, elles ne sont plus obligées de déclarer leur pathologie après l'écoulement de certains délais fixés d'1 à 10 ans selon le type d'affection et intégrés dans une grille de référence adoptée en février 2016. En l'absence de rechute, la fin du protocole thérapeutique est le point de départ du délai décennal. - Cf. de Fallois M., Assurance et " droit à l'oubli " en matière de santé, RDSS 2017-1, p. 132. - Bouteille-Brigant M., Les indispensables du droit médical, Ellipses, 2016, p. 98.

formations sur leur santé, pour bénéficier de l'assurance, sans surprime ni exclusion. Concrètement, lors de la souscription d'un contrat assurance emprunteur, ils ne sont plus tenus de déclarer leur pathologie après l'écoulement de certains délais fixés d'un à dix ans selon les six types d'affection (hépatite virale C, cancer du testicule, cancer de la thyroïde, certains cancers du sein, mélanome de la peau et cancer du col de l'utérus) et intégrés dans une grille de référence adoptée le 4 février 2016. En l'absence de rechute, la fin du protocole thérapeutique est le point de départ du délai décennal. Regrettablement, les assureurs demeurent libres de refuser de garantir des personnes présentant un risque aggravé de santé candidats à des prêts immobiliers et professionnels. La convention AERAS ne leur garantit aucun accès au crédit. Par ailleurs, les assureurs sont susceptibles d'utiliser de nouvelles technologies, comme la blockchain, dans sa forme de registre sécurisé par exemple, laquelle ne permettrait pas d'effacer les données personnelles. L'assureur peut ainsi trouver un fort intérêt à connaître l'historique d'un contrat (sa reconduction, ses avenants, pour l'application de la garantie dans le temps) et l'assuré à savoir si une clause, abusive par exemple, a bien été éradiquée de la police. Le problème surgit en matière de données protégées, pêle-mêle par la confidentialité, le secret professionnel, le secret médical, ou en matière de données sacrifiées sur l'autel de la transparence par le droit à l'oubli. Deux formes récentes de l'oubli ont été érigées en droit subjectif par le législateur : la convention AERAS pour les assurances des emprunteurs ayant eu des maladies graves et encore l'appréciation par le procureur de la République de l'opportunité de l'effacement dans le fichier de traitement d'antécédents judiciaires pour des personnes ayant eu un tel passif. Le risque est donc celui d'une violation ineffaçable dans la blockchain, au point qu'elle puisse devenir perpétuelle. Reportée sur l'analyse du risque par l'assureur, cette impossible suppression de l'information dans la blockchain est susceptible de construire un système d'entrave anticipée à la réalisation du droit à l'oubli.

Question à l'actuaire :

L'actuaire est-il déjà amené à utiliser des blockchains dans l'assurance ? Le cas échéant, ces registres dématérialisés posent-ils des difficultés de collecte et de traitement des données à l'actuaire eu égard à certaines informations sensibles, éventuellement protégées par la confidentialité ou le secret, et susceptibles d'influer sur la nature du risque ? En matière de santé concrètement, le croisement et la connaissance de données passées ne permettent-ils pas de contourner le droit à l'oubli et de refuser, en toute connaissance du risque, un assuré qui devrait être considéré comme vierge par suite du droit qui lui est reconnu de taire sa pathologie ? L'actuaire peut-il aussi identifier des données protégées par exemple par le droit à l'oubli et les mettre à l'écart pour ne pas discriminer ni à l'entrée ni à la tarification ?

AC : Les blockchains sont aujourd'hui utilisées en assurance pour éviter de passer par des intermédiaires coûteux, et n'apportant pas grand-chose dans la chaîne de valeur. Certains assureurs proposent déjà des contrats pour des retards relatifs à des voyages (avion ou train). Historiquement, l'assuré devait contacter son assureur afin de déclarer un retard, avec des justificatifs (parfois difficiles à obtenir). L'assureur devait ensuite vérifier l'information, puis procéder à l'indemnisation. Ces contrats sont simples car l'indemnité est directement liées à une information qu'il est possible d'avoir par ailleurs, de manière sécurisée (il existe des registres attestant des heures de décollages et d'atterrissages). L'idée des contrats de type blockchain est de proposer une assurance indicielle ou paramétrique qui ne nécessite plus de demande de la part de l'assuré et de validation par l'assureur. Le processus peut être automatisé et sécurisé. On peut imaginer des contrats indiciels agricoles basés sur la même technique, avec le versement (automatique) d'une indemnité s'il n'a pas plu pendant trente jours consécutifs. Pour l'instant, les blockchains ne posent pas de soucis quant aux données utilisées car elles sont souvent publiques (heures d'arrivée d'avion, indice de température). Mais la question se poserait pour des contrats dont le sous-jacent serait une donnée personnelle. Pour le second point, je n'ai pas beaucoup d'expérience en données de santé. Néanmoins, prenons le cas de l'assurance automobile. Quand on parle ici d'une utilisation de données passées, on parle de personnes déjà présentes depuis plusieurs années dans le portefeuille, souhaitant un renouvellement. Dans les exemples mentionnés, on parle plus précisément d'informations relatives à des sinistres : s'il a eu une condamnation pour excès de vitesse, l'assureur ne le sait pas, sauf s'il y a eu un accident. Or le droit des assurances donne déjà un pouvoir discrétionnaire à l'assureur d'exclure certains assurés suite à un sinistre, donc je ne suis pas sûr qu'il y ait quelque chose de réellement nouveau. La difficulté se pose quand l'information que l'assureur veut utiliser n'est pas une information relative à la triche (ou une fraude) avérée d'un assureur, mais sur la suspicion de fraude. Supposons qu'un gestionnaire de sinistre ait la possibilité d'indiquer sur certains dossiers « suspicion de fraude » (et que cela engage ensuite l'envoi d'un expert). Pourrait-on utiliser cette information même plusieurs années après, ou peut-on imaginer un droit à l'effacement ? Dans ce cas, les techniques statistiques viennent sauver les assurés. En effet, si tous les sinistres étaient contrôlés par des experts, il n'y aurait plus de « suspicion » de fraude mais juste une variable « a fraudé » / « n'a pas fraudé ». Mais c'est plus complexe car sur le cas de la fraude, l'actuaire dispose juste d'un échantillon, d'une sous-population sur laquelle un expert s'est prononcé sur la fraude. Si l'envoi d'un expert est purement aléatoire, on aura un sondage sur une population

représentative. Mais ce n'est souvent pas le cas : l'expert est souvent envoyé suite à une suspicion. La sous-population n'est pas représentative, et il est alors très difficile d'utiliser la variable fraude. Cette variable est dite manquante car on ne sait pas si la personne a fraudé ou pas, on sait juste que la personne n'a pas été jugée suspecte. Ce biais rend l'utilisation de ces données très délicates. C'est en fait pareil pour l'assurance santé dont on parlait : que signifie le fait qu'aucune maladie n'ait-été déclarée ? Qu'il n'y a pas eu de maladie, ou bien qu'elle n'a pas été mentionnée ? Travailler sur des données dites manquantes rend l'exercice délicat car l'interprétation peut être fallacieuse.

Des données dont la pertinence est sujette à évolution.

DCS : L'on imagine volontiers que le travail d'actuariat évolue, selon les transformations sociales, environnementales, les techniques disponibles… Si l'âge, le lieu de résidence ou l'activité professionnelle sont des facteurs de discrimination tarifaire notoirement connus, ils relèvent d'une modélisation classique, liée à une époque où notamment les sources et les méthodes de traitement des données était moins développées qu'aujourd'hui…

Question à l'actuaire :

Y a-t-il une limite aux types de données dont vous êtes susceptible de découvrir la pertinence pour mesurer le risque en assurance de masse (automobile, multirisque habitation, santé) ? L'actuariat va-t-il dégager de nouveaux critères, s'ils ne le sont déjà, notamment dans un contexte de croissance du big data ?

AC : Historiquement, ces variables tarifaires étaient utilisées car elles étaient faciles à avoir. Il y en a qui seraient intéressantes, mais plus difficiles à avoir. Par exemple a-t-on un gros conducteur ou un conducteur occasionnel ? Le kilométrage est une variable très liée à la sinistralité. Mais elle est inconnue à la souscription. Une astuce a longtemps été de demander à ce que l'assuré s'engage à faire moins de 7000 km par an (par exemple). Autre information longtemps utilisée, le genre. Là encore, c'était une variable facile à obtenir, et corrélée avec la sinistralité. Les études récentes sur données télématiques ont montré que le genre était utilisé comme le « proxy » d'une information difficile à avoir, sur le type de trajets effectués. En particulier, en utilisant des informations sur l'heure à laquelle les trajets sont effectués, le nombre de kilomètres parcourus, etc., le genre n'apportait aucune information supplémentaire intéressante[13].En santé, on pouvait demander si la personne fumait ou pas. Mais ça reste déclaratif, et difficilement vérifiable. On pourrait imaginer avoir des informations plus précises par des objets connectés. Ou sur la pratique sportive d'une personne (nombre de pas par jour avec des bracelets connectés).

Le cas des données recueillies via des objets connectés

DCS : Vous venez d'évoquer les objets connectés, et l'on voit précisément à cet égard de plus en plus d'initiatives tendant à lier le contrat d'assurance à leur utilisation. Ainsi en matière automobile, les contrats Youdrive de Direct Assurance et Rate my drive d'Aviva GB, ou en matière de santé le contrat Vitaly proposé par Generali Assurance. Ces contrats restent pour l'instant des alternatives aux contrats classiques en raison, semble-t-il, de la méfiance naturelle de la majorité des assurés, soucieux de leur vie privée, et reposent donc sur l'adhésion à ce type de procédé. Ces nouvelles pratiques soulèvent donc au premier chef la question de leur dépendance au consentement de l'assuré. S'y ajoute celle du moment de la collecte des données de l'objet connecté, notamment au regard de sa finalité.

La question du consentement de l'assuré

A l'heure actuelle, l'assuré consent à l'utilisation d'objet connecté afin de bénéficier d'un tarif mieux ajusté (tarif bon conducteur), voire d'avantages en nature (assurance santé). A côté du potentiel de prévention que vous évoquiez précédemment, l'on pressent certainement aussi la possibilité d'analyser plus finement le risque grâce aux nouvelles données collectées par l'objet. Or l'on a rappelé le lien existant entre l'entrée d'une donnée dans le champ de l'évaluation du risque et l'obligation pour le candidat à l'assurance de la livrer. Et si cette pratique ne semble être fondée à l'heure actuelle que sur le consentement de l'assuré, il n'est pas interdit de se demander ce qu'il en serait dans le cas où la pression concurrentielle conduirait à généraliser ce type de contrats.

Une pratique supposant le consentement de l'assuré

Le recours à l'objet connecté, qu'il s'agisse de prévenir ou d'évaluer le risque, est aujourd'hui fondé l'adhésion de l'assuré et son utilisation doit donc consentie. Mais à défaut d'une information précise sur le type d'informations collectées, il est difficile de considérer que le consentement à l'utilisation du dispositif (boitier sur véhicule, indicateur d'activité physique pour santé) emporte ipso facto consentement au traitement de toute donnée collectée par son biais. Un tel raccourci serait contraire à l'exigence d'un consentement spécialement donné que le Règlement consacre[14]

[13]Résultat établi par R. Verbelen, K. Antonio et G. Claeskens dans Unraveling the Predictive Power of Telematics Data in Car Insurance Pricing (http://bit.ly/2DBNMto).

[14]On rappellera ici que l'exigence du consentement a été considérablement renforcée dans le texte européen. Ainsi notamment, l'article 7.2 dispose-t-il que « si le consentement de la personne concernée est donné dans le cadre d'une déclaration écrite qui concerne également d'autres questions, la demande de consentement est présentée sous une forme qui la distingue clairement de ces

lorsque le traitement repose sur ce fondement. Par ailleurs, le droit à l'autodétermination informationnelle[15] permet à la personne concernée de retirer celui-ci (art. 7.3 RGPD). Techniquement cela devrait au moins se traduire par une maitrise directe de l'assuré sur l'ensemble du flux de données (ex. V. le nouveau « Pack Assurance » de la CNIL qui envisage le cas du véhicule connecté transmettant les données – V. spéc. le scenario in-out– et suggère de prévoir des réglages par défaut sans transmission, ou des dispositifs simples de coupure du flux). Cependant, pour que le droit d'autodétermination informationnel soit effectif, ou simplement même pour que l'on puisse parler de consentement au traitement, cela suppose que l'assuré soit conscient de la nature des données que l'objet collecte, et plus avant, de celles que leur croisement est susceptible de fournir à l'assureur. En d'autres termes, analyser la portée du consentement donné par l'assuré à l'usage de l'objet connecté dans le cadre de son contrat d'assurance suppose de se pencher plus avant sur le type de données dont l'objet connecté sera la source et sur l'usage que peut en faire l'assureur.

Questions à l'actuaire :

Pouvez-vous nous donner une idée du champ de données que ce type d'objets (boitier véhicule, bracelet porté par l'assuré) est techniquement en mesure de fournir à l'assureur ? Au-delà des données brutes (géolocalisation par ex.), quelles autres informations (et donc nouvelles données) l'assureur est-il susceptible de recueillir à partir de leur croisement ?

AC: Pour les données télématiques, les données fines permettent d'avoir des scores de freinage, d'accélération, qui donne des informations sur le type de conduite. Techniquement, ils permettent de savoir aussi qui conduit (voiture conduite par monsieur et madame, voire le grand enfant dans le cas de la conduite accompagnée). On peut aussi savoir quel type de route est utilisé, quelle distance est parcourue par jour, etc. Mais il est aussi possible d'avoir des informations plus agrégées, comme le pourcentage de temps de conduite la nuit. La difficulté rejoint une précédente question : les données collectées dépendent du type de boitier utilisé, mais pour savoir ce qui serait utile, il faut avoir collecté les données. Les assureurs font généralement toutes sortes de tests, afin de savoir ce qu'ils exploiteraient réellement, et à quelle fin. Et voir s'ils veulent les données en temps réel, ou juste à des fins statistiques, en fin de semaine ou de mois. Un assureur pourrait être intéressé par ces données en temps réelles pour proposer une application liée à la prévention du risque, par exemple offrir un café sur une aire d'autoroute si le conducteur a déjà conduit deux heures d'affilé, et qu'il s'engage à faire une pause de trente minutes.

L'hypothèse d'une collecte imposée.

DCS : Si la pratique actuelle des objets connectés dans les contrats d'assurance repose sur le consentement de l'assuré, les raisons en sont néanmoins pour l'instant essentiellement commerciales. En effet, les assurés ne sont pas prêts à les admettre si facilement les potentielles intrusions des objets connectés dans leur vie privée. La question se pose néanmoins de savoir si ce consentement est, du point de vue juridique, un rempart absolu contre ces pratiques. Ainsi qu'on l'a vu, l'assuré ne peut refuser, au sens du Code des assurances, de livrer « les circonstances qui sont de nature à faire apprécier par l'assureur les risques qu'il prend en charge » (C.ass., art. L. 113-2). Pour l'instant cette obligation est, de fait, circonscrite au cas où l'information est sollicitée via le questionnaire d'assurance, mais elle n'est pas, en droit, expressément limitée à cette forme de collecte[16]. Or, il n'est pas interdit d'imaginer qu'une information, que seuls ces objets peuvent livrer (ex. données fines de trajet) puisse un jour être tenue pour une circonstance « de nature à faire apprécier le risque par l'assureur ». Dans un tel cas, l'obligation de livrer l'information qui va de pair avec cette qualification pourrait se transformer en une obligation d'accepter l'usage de l'objet connecté pour celui qui veut s'assurer. Et, du point de vue, plus général, adopté par le Règlement, l'assureur serait fondé à récupérer cette donnée dès lors que cela est nécessaire à l'exécution du contrat, autrement dit sans avoir à recueillir le consentement de la personne concernée (soit qu'elle conditionne la souscription, soit qu'elle soit faite éventuellement à son insu).

Question à l'actuaire :

Pensez-vous disposer un jour de suffisamment de ces « nouvelles données », issues d'objets connectés, pour que la modélisation du risque repose dessus, au point que l'assureur qui aura construit son modèle tarifaire sur ces variables ne veuille – voire ne puisse - plus s'en passer pour l'ensemble des assurés ? La conséquence n'en serait-elle pas, à terme, une généralisation de la pratique des assurances liées à des objets connectés du fait de leur intégration dans le modèle tarifaire (art. L. 113-2, 2°) ?

AC : Un tarif qui utiliserait ces « nouvelles données » pose la

autres questions, sous une forme compréhensible et aisément accessible, et formulée en des termes clairs et simples ». Le consentement au traitement des données doit donc être spécialement donné.

[15] Introduit matériellement en droit français par loi n 2016-1321 du 7 octobre 2016 pour une République numérique, qui a ajouté à l'article premier de la loi I&L un nouvel alinéa selon lequel « toute personne dispose du droit de décider et de contrôler les usages qui sont faits des données à caractère personnel la concernant, dans les conditions fixées par la présente loi ».

[16] V. la rédaction de l'article L. 113-2 qui évoque les questions posées à l'assuré « notamment dans le questionnaire d'assurance ».

question importante de l'absence d'information : que faire si une personne refuse qu'un boitier GPS soit installé en permanence ? Il existe plusieurs solutions. La première est d'imposer le boitier mais qu'une désactivation soit possible. Les assureurs qui ont essayé se sont rendu compte qu'il y avait alors un biais important dans les données collectées (trop peu utilisaient leur voiture du vendredi après-midi au dimanche matin, au vu des données). Une seconde est de faire deux contrats, un pour les utilisateurs de boitiers, un pour les non-utilisateurs. A l'heure actuelle, les assureurs cherchent à avoir des données, le plus possible, afin de mieux comprendre quelles informations sont réellement pertinentes dans un modèle prédictif. Il y a une tendant à offrir un « bonus » aux personnes qui acceptent de fournir leurs données. Mais si l'installation de boitier n'a pas d'impact sur la sinistralité, c'est un jeu à somme nulle (au sens où le montant des sinistres à payer est inchangé) : un « bonus » pour une sous-population signifie un « malus » pour l'autre. Autrement dit, il y aura une pénalisation pour les personnes qui ne souhaitent pas céder ces données personnelles. Cela n'est pas sans poser des problèmes à l'actuaire, car ce refus est une variable intéressante, importante, dont les liens avec la sinistralité sont difficiles à appréhender. Il convient peut-être de rappeler ici qu'il existe deux tarifications relativement différentes. La première est une tarification dite « a priori », qui concerne les nouveaux clients ; la seconde est une tarification dite « a posteriori » (on parlera parfois d'« experience rating »), qui concerne les renouvellements des contrats. Dans les cas évoqués ici, l'information est alors connue une fois le contrat souscrit et on ne peut pas utiliser ces données en tarification (à moins de changer la forme des contrats. Par exemple en offrant des contrats d'assurance automobile non plus à l'année, mais pour un nombre de kilomètres prédéterminés). En revanche, il serait possible d'utiliser cette information lors d'un renouvellement. Par exemple on pourrait dire que l'an passé, le conducteur a fait partie des 10% des plus gros conducteurs (en terme de nombre de kilomètres parcourus) et qu'une majoration de prime s'impose.

La question du moment de la collecte.

RB : En définitive, un décalage temporel apparaît entre le moment où les données sont mises à disposition et celui de la conclusion du contrat d'assurance. Les objets connectés accouplés à des smart contract seraient parfois susceptibles de surmonter ce décalage. Il subsiste un problème intermédiaire, celui du déclenchement de la garantie (d'une assurance automobile) en fonction d'un objet connecté relié à un smart contract. On peut imaginer deux modes de consentement : une souscription à l'assurance en amont dont le smart contact est une condition liée à une prise d'information en temps réel (condition suspensive de mise en œuvre de la couverture liée au démarrage du véhicule) ou une souscription à l'assurance in situ, à chaque évaluation du risque par l'objet connecté. Outre les problèmes du consentement indispensable en principe pour la formation du contrat, la prise de données de l'assuré et des biens qu'il assure en permanence par l'assureur peut éventuellement lui permettre de les utiliser ultérieurement à son encontre, c'est-à-dire à d'autres fins que l'objet et l'obligation pour lesquels et dans le cadre desquels l'assuré a consenti à l'installation de l'objet connecté.

L'exploitation des données de l'assuré par l'assureur

DCS : Une fois collectées, les données d'assurés sont susceptibles d'être exploitées de différentes manières par l'assureur. Au premier chef, bien sûr, lors de la souscription du contrat, s'agissant d'accepter ou non le risque, et le cas échéant, de le tarifer. Au cours de l'exécution du contrat, la question de l'usage qui est fait des données de l'assuré ressurgit dans le cadre de la lutte contre la fraude. Le cycle des données personnelles d'un assuré ne s'achève cependant pas avec l'exécution du contrat, car celles sont encore susceptibles d'être exploitées parfois bien au-delà du terme du contrat, du fait qu'elles ont vocation à intégrer les statistiques actuarielles de l'assureur.

Décision de prise en charge et tarification du risque

En raison des spécificités de l'opération d'assurance, l'exploitation des données d'assuré pour déterminer la prise en charge du risque ou sa tarification est en premier lieu confrontée à l'appréhension par le droit de certains types de discriminations, soit qu'il les interdise, soit qu'il cherche simplement à les réguler. S'agissant plus précisément de la tarification, les techniques mises en œuvre dans ce cadre invitent à se pencher plus spécialement sur la question du profilage.

Discrimination prohibée ou limitée. *Discrimination tarifaire prohibée*

Depuis le 1er mars 2011, date de l'arrêt Test Achat rendu par la CJUE (Affaire C-236/09) sur le fondement de la Directive 2004/113 du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans la fourniture de biens et services, la discrimination fondée sur le sexe de l'assuré est clairement condamnée. Or ce critère faisait partie des facteurs actuariels historiquement utilisés par les assureurs pour prendre la mesure des risques (différenciation de la longévité, du risque automobile…). L'assureur, dont on rappelle qu'il cherche à limiter les effets de l'antisélection, doit donc se fonder sur d'autres facteurs. Peut-être est-il alors tenté d'en chercher de nouveaux. Sachant que l'on peut a priori déduire des sympathies politiques du simple croisement de données sur les préférences musicales, il n'est pas interdit de penser que le sexe puisse être déduit du croisement d'autres données, et de s'interroger sur le caractère très platonique de telles interdictions à l'ère du big data.

Question à l'actuaire :

Comment les assureurs ont-ils intégré la disparition d'un facteur discriminant majeur de leurs modèles d'évaluation du risque après 2011 ? Techniquement, le fait de disposer de multiples données périphériques ne permet-il pas de déduire la donnée discriminante « interdite » et donc de masquer le réel critère de discrimination tarifaire ? Est-il possible que l'assureur, même sans chercher délibérément à « reconstruire » la donnée « interdite » (sexe dans notre exemple), retrouve la même granularité dans l'analyse du risque à partir des nouvelles données ?

AC: Comme nous l'avons évoqué, il convient de repenser de manière historique comment les tarifs ont été construits. Si l'âge était utilisé en assurance automobile, c'est que cette variable est facile à obtenir, et qu'elle est corrélée à la sinistralité. Certains évoquent aujourd'hui l'idée de « spurrious correlation » dans le sens où aucune relation causale ne peut être établie, et elle serait un proxy d'une variable plus difficile, historiquement, à observer. Beaucoup évoquent l'idée que l'âge devrait aussi être exclu, tout comme le genre. En assurance automobile, l'expérience est un facteur clé. Une parade serait d'utiliser l'ancienneté du permis, qui est en fait, bien souvent, une variable plus informative que l'âge (même s'il existe des particularités pour les personnes d'âge très avancé). Il faut garder à l'esprit la différence entre « avoir accès à certaines données » et « pouvoir avoir accès ». Sur le genre, on peut imaginer que l'actuaire puisse prédire (avec un faible taux d'erreur) le genre du conducteur. Mais pour ce faire, il utilisera probablement des données plus intéressantes pour la tarification. Chercher à recréer la variable de genre serait alors un travail coûteux, et inutile.

La discrimination limitée

RB : S'agissant de l'exploitation des données sensibles de l'assuré par l'assureur, de santé tout d'abord, ce dernier a la faculté de questionner le futur assuré sur sa santé et sélectionner le risque sans encourir le grief de discrimination et sans violer l'interdiction du traitement de telles données telle que formalisée par l'article 9 du RGPD (art. 8 Loi Informatique et Libertés de 1978). En France, les tests génétiques demeurent interdits néanmoins. L'assuré est ainsi amené à délivrer des informations que l'on peut qualifier, souvent, de données très personnelles, sensibles même. L'objectif est d'offrir un tarif adapté, au plus grand nombre, comme offre d'assurance en contrepartie. Le droit à l'assurance peut alors ne pas être en totale adéquation avec l'obligation d'assurance. La situation paraît plus encore difficultueuse en présence de risques aggravés de santé. Certes, la convention AERAS, s'assurer et emprunter avec un risque aggravé de santé, et ses avenants, ont pour objet de prendre en compte ces risques spécifiques, avec un aménagement spécial des questions de discrimination. Il s'agit donc d'affiner la sélection du risque particulier pour offrir une couverture assurantielle dans certaines limites, grâce à l'amélioration du profil du risque particulier d'une part et de la connaissance des pathologies en général d'autre part. Il ne semble pas être recherché la capacité à l'assurabilité de tout le monde indifféremment.

Question à l'actuaire :

Comment l'actuaire prend-il en compte le nouveau dispositif AERAS dans sa modélisation actuarielle ?

AC : le but d'un modèle actuariel (qu'il s'agisse d'une étude statistique afin de mieux comprendre le risque, ou de la constitution d'un tarif) est de trouver des variables qui pourraient être corrélées à la survenance d'un risque, ou son coût. Dans le dispositif AERAS, un assuré a le droit de ne pas déclarer un cancer passé, dont le protocole thérapeutique est terminé depuis 10 ans, sans rechute. C'est ce que dit la loi. La question importante est de comprendre comment cette loi a été élaborée, et en particulier pourquoi cette clause a été proposée. Si elle est motivée par des études épidémiologiques, qui montrent que ces personnes ont le même risque de redévelopper un cancer qu'une personne n'ayant jamais eu de cancer diagnostiqué, alors l'actuaire n'a pas de motivation pour essayer de retrouver cette information, car elle n'apporte rien. En statistique, c'est la notion d'indépendance entre des variables (la non-corrélation est une version un peu plus simple) : les évènements A et B sont indépendants si le fait de savoir B n'apporte aucune information sur A. Savoir qu'il pleut à Tokyo ne m'apporte aucune information sur le fait qu'il pleuve à Paris, donc si je fais un modèle météo, je n'ai aucun intérêt à avoir accès à cette information. Ici aussi, si le fait d'avoir eu un cancer il y a 10 ans (et d'avoir survécu, sans rechute) n'influence pas la probabilité d'avoir un cancer aujourd'hui, il n'y a aucune raison de chercher à avoir cette information. Il faut garder en mémoire que les actuaires et les statisticiens qui font de l'épidémiologie ont été souvent formés ensemble, utilisent les mêmes modèles (voire les mêmes logiciels). La situation serait toutefois différente si cette loi avait été votée sur d'autres bases. Si pour des raisons politiques (électoralistes ou éthiques, peu importe – du point de vue du statisticien) il a été décidé de rajouter cette clause, alors que le risque est inchangé, ce n'est pas la même chose. On revient ici sur un problème bien connu en économie d'asymétrie d'information (entre l'assureur et l'assuré), où la solution est de faire révéler l'information (par un moyen détourné) à l'agent le plus informé.

L'automatisation des processus décisionnels (profilage) .

DCS : Il a été rappelé que le domaine des assurances est intrinsèquement associé à la segmentation tarifaire au sein

de la mutualité. Qui dit segmentation dit classement des risques pris en charge en fonctions de critères prédéfinis. Ce mécanisme est par essence de ceux qui bénéficient du développement des procédés modernes d'automatisation des traitements. On parle alors de profilage. Le RGPD évoque sous ce terme « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ».

Dès lors que la conclusion du contrat d'assurance est automatisée (ce qui devrait être de plus en plus fréquent, si l'on pense notamment au développement des assurances en ligne), le rattachement de l'assuré à l'une des catégories tarifaires de l'assureur en fonction de son niveau de risque ne peut mieux répondre à une telle définition. Il semble bien s'agir du profilage visé tant par le Règlement que, avant lui, par la loi Informatique et libertés de 1978. Mais encore faut-il que le scoring conduise à une décision « produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire » (élément définitoire du « profilage »), ce qui, si l'automatisation permet d'accepter (ou de refuser) et de tarifer la couverture proposée, ne saurait faire de doute.

Le profilage est interdit en matière de données dites sensibles[17] et le Règlement rappelle le droit des personnes concernées de « ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé », mais il en va en revanche autrement dès lors que « la décision [...] est nécessaire à la conclusion [...] d'un contrat » (art. 22.2, a). Si bien que celui qui demande à être assuré ne saurait a priori s'opposer au profilage qui serait pratiqué par l'assureur à la souscription. L'assuré se voit néanmoins conférer deux droits spécifiquement associés à une telle hypothèse. L'un concerne le droit d'être informé sur l'existence et la logique sous-jacente du profilage. L'autre auquel nous nous intéresserons en premier lieu, serait celui de pouvoir réintroduire une intervention humaine dans le processus décisionnel.

La question du droit à une intervention humaine

Question à l'actuaire :

L'actuaire dispose-t-il des moyens d'intégrer cette éventualité dans les modèles d'évaluation du risque ? Autrement dit, est-il possible d'élaborer un modèle laissant place le cas échéant à une intervention humaine ? Quels pourraient-être les critères auxquels aurait recours l'intervenant humain et qui n'auraient pas déjà été intégrés dans le processus décisionnel par l'actuaire ? Peut-on selon vous réellement distinguer, dans la pratique assurantielle, des décisions automatisées et des décisions impliquant une « intervention humaine » ? Enfin, est-il envisageable que le « point de vue » d'un assuré qui contesterait la décision tarifaire (ou d'exclusion de la couverture) puisse changer la décision qui sera prise à son égard par l'assureur ?

AC : Pour l'instant, les modèles sont très artisanaux, dans le sens parfois entendu où l'actuariat est à la fois un art et une science. L'actuaire choisit quelle variable utiliser, quelle technique permettra de lisser le zonier, quelles interactions sont pertinentes, etc. Mais le « machine learning » (et plus généralement toutes les techniques liées à l'intelligence artificielle) propose aujourd'hui des techniques (facilement programmables) qui permet de ne faire (presque) aucun choix. Historiquement, on avait des modèles très simples, avec des classes de risques très clairement identifiées : une prime de base, un rabais de 15% pour les femmes, une hausse de tarif de 20% pour les personnes habitants en banlieue, et de 30% pour les jeunes conducteurs, puis on rajoute une majoration de 15% pour un véhicule diesel, etc. Puis les modèles sont devenus plus complexes, avec la prise en compte de variables « continues » (l'âge et non plus la classe d'âge, ce qui permet de lisser une baisse ou une hausse tarifaire) mais aussi des croisements de variables (avec une pénalisation pour une classe d'âge spécifique, une zone géographique et un type de véhicule précis, par exemple). Si les modèles étaient plus complexes, les actuaires maitrisaient toujours leur construction. Plus récemment, les actuaires ont découvert l'enrichissement de données, correspondant à lier les données possédées par l'assureur à des bases plus importantes (données sur les caractéristiques du véhicule, sur les caractéristiques du logement, des statistiques sur le quartier où réside la personne, sur les garages à proximité, etc.), il devient impossible de tester la pertinence de toutes les variables. Ce problème de grande dimension a été en partie résolu par les techniques de « machine learning » qui proposent une sélection automatique des variables. Il n'est pas étonnant de voir Google ou Amazon intéressés pour entrer sur le marché de l'assurance. Les techniques actuelles sont proches de celles qu'ils développent.

Au-delà du modèle, le tarif repose surtout fondamentalement sur les données. Le même modèle (même le plus simpliste) donnera des primes différentes sur des populations différentes. Le même modèle sur le portefeuille d'assurés d'AXA ou la mutualité de sociétaires de la MAIF donneront des primes (ou des cotisations) très différentes. Contester un prix signifierait qu'il pourrait exister « un prix » à un risque donné. Or rien n'est plus faux. Un risque en assurance ne

[17] Sauf consentement de l'intéressé et motif d'intérêt public, validé par le droit d'un État membre (art. 22.4 renvoyant aux exceptions de l'art. 9.2, a ou g)

peut être vu qu'au sein d'un groupe qui cherche à mutualiser le risque, au sein d'une compagnie. La valorisation en assurance n'est pas celle des marchés financiers, où le principe de réplication prévaut. En mathématique financière, deux actifs qui rapportent la même chose dans tous les états de la nature ont forcément le même prix (sinon via la vente à découvert, il serait possible de gagner de l'argent sans prendre de risque, et de réaliser un « arbitrage »). En assurance c'est impossible. Et il est tout à fait possible d'être vu comme un « bon » risque pour tel assureur, et comme un « mauvais » pour tel autre.

Cela dit, les actuaires proposent un calcul d'une prime dite technique. C'est la prime qu'il conviendrait de demander pour avoir un portefeuille équilibré, et concurrentiel. Mais c'est rarement la prime réellement payée par l'assuré, car il convient d'intégrer une stratégie commerciale, mais aussi le fait que l'activité d'assurance fait intervenir bon nombre d'intermédiaires (entre les agents et les courtiers).

La question du droit à l'information sur la logique du profilage

DCS : Le second droit que le RGPD confère à l'occasion de la mise en œuvre d'un profilage, relève plutôt de l'idée de transparence. Ainsi l'article 13 prévoit-il que le droit à l'information de la personne concernée porte notamment sur « l'existence d'une prise de décision automatisée, y compris un profilage, [...] et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée » (art. 13.2, f).

De fait, cela confère potentiellement à tout assuré le droit de connaitre, et à tout assureur l'obligation de notifier, les critères utilisés par le second pour placer le premier dans telle ou telle catégorie tarifaire, voire si l'on se réfère à la notion de « logique sous-jacente » peut-être même à l'algorithme utilisé pour évaluer son risque.

Question à l'actuaire :

Comment l'assureur pourrait-il satisfaire en pratique à une telle obligation d'information ? Les algorithmes utilisés pour les décisions de tarification ne sont-ils pas des secrets commerciaux ? Et si l'on ne révèle pas l'algorithme, que pourraient être ces « informations utiles concernant la logique sous-jacente » du profilage ? Autrement dit, les algorithmes utilisés comme supports du contrat d'assurance sont-ils susceptibles d'être « résumés » ou vulgarisés ?

AC : Le problème est compliqué. La fonction de l'actuaire est de faire du profilage. C'est aujourd'hui la base de l'assurance. Les algorithmes sont des « secrets commerciaux », mais peut-être moins que les données. En réalité, tous utilisent les mêmes outils, les mêmes techniques. Mais comme ils ont des portefeuilles différents, ils ont des paramètres différents. Par exemple des mutuelles comme la MATMUT ou la MAIF, qui historiquement assurent des fonctionnaires, avaient des portefeuilles très différents d'assureurs privés. Même en construisant le même modèle, les primes demandées sont très différentes. Techniquement, si tous les assureurs expliquaient quelles variables ils utilisent, ça ne changerait pas grand-chose. Fondamentalement, les assureurs connaissent structurellement leur modèle de tarif. Une autre difficulté est de comprendre ce que veut dire « utiliser une variable ». Oui, un assureur va utiliser le modèle du véhicule, parce qu'il peut ensuite obtenir sa côte à l'argus, parce qu'il va en déduire son poids, sa vitesse maximale, sa puissance. Avec le croisement des données, il est complexe d'expliquer lors de la signature du contrat comment une information est utilisée, et si elle l'est. De plus, cet enrichissement de données peut donner lieu à de très légers changements, insignifiants, infinitésimaux mais significatifs. Et qui mis bout à bout ont des conséquences importantes. Cela dit, la règle peut avoir des conséquences intéressantes : elle permettrait que l'assuré puisse agir sur les variables qui le classent comme un « risque élevé » en demandant une prime plus importante que la moyenne. C'est intéressant car cela permet de mettre en place de la prévention. Si l'assuré sait que sa prime d'assurance multirisques habitation est élevée c'est parce qu'il n'a qu'un verrou sur sa porte d'entrée, il a la possibilité d'installer un second verrou, ce qui baissera la prime, mais aussi le risque. En revanche, s'il sait que sa prime d'assurance automobile dépend de la puissance de son véhicule ou du type de carburant, il ne va pas revendre son véhicule pour un acheter un autre. Et souvent, ces variables sont croisées avec d'autres : le carburant ne joue plus via un simple coefficient majorateur (par exemple +20% pour un véhicule diesel) mais le carburant peut-être croisé avec l'âge du véhicule, et le lieu d'habitation. Donc avoir cette information ne servirait pas à grand-chose en pratique.

La lutte contre la fraude

DCS : Le fondement d'un traitement de données personnelles dans le cadre de la lutte contre la fraude à l'assurance ne pose pas de problème particulier. Il relève à n'en pas douter de l'intérêt légitime de l'assureur, au sens de l'article 6.1.f du RGPD qui confirme en substance le régime antérieur du droit français[18]. La « prévention de la fraude » est d'ailleurs visée au premier chef dans le considérant 47

[18] V. spéc. sur ce point l'autorisation unique (AU) élaborée par la CNIL : Délibération n 2014-312 du 17 juillet 2014 portant autorisation unique de traitements de données à caractère personnel ayant pour finalité la lutte contre la fraude à l'assurance mis en œuvre par les organismes d'assurance, de capitalisation, de réassurance, d'assistance et par les intermédiaires d'assurance (AU 039)

comme illustrant un intérêt légitime du responsable de traitement. A ce titre, seront licites tant l'exploitation des données initialement collectées pour une autre fin (par ex. pour l'appréciation du risque dans le questionnaire d'assurance), que les nouvelles collectes de données susceptibles d'établir la fraude (ex. activités de l'assuré depuis le sinistre).

Si le traitement sera donc a priori licite, la recherche des cas fraudes est néanmoins susceptible de générer des pratiques qu'il convient de confronter au droit spécial des données personnelles, voire à d'autres règlementations intimement liées à la protection de la vie privée. Ainsi, après le processus de sélection et de tarification des risques, la question du profilage se pose de nouveau en matière de détection de la fraude. On retrouve par ailleurs naturellement certaines difficultés liées à la question de la loyauté de la preuve.

La question du profilage des fraudeurs. Les analyses prédictives ne se cantonnent pas au domaine du risque et se retrouvent parfois également dans le cadre de la lutte anti-fraude. Ainsi Aviva Assurances a démarré un projet de croisement numérique de ses dossiers internes de fraudes afin de créer des modèles par corrélation de cas avérés et compte ainsi détecter les 4/5e du potentiel de cas qui échapperaient encore à sa cellule anti-fraude[19]. L'utilisation de scores prédictifs (scoring de suspicion des sinistres) permettrait ainsi de décupler les potentialités en matière de traque à la tricherie. Si ce type de profilage (au sens commun du terme) semble pour l'instant à un stade expérimental, il n'en relève pas moins du droit de la protection des données sur lesquelles il s'appuie.

Or, la question ne peut être envisagée ici de la même manière qu'à la conclusion du contrat. En effet, contrairement à ce qui se passe lors de l'évaluation du risque, il ne devrait pas y avoir par hypothèse de décisions basées sur un traitement automatisé, ce qui définit le profilage au sens du RGPD. Si l'assureur peut s'appuyer sur des logiques prédictives et des processus automatisés pour déceler plus facilement les cas de fraude, il ne saurait a priori refuser d'indemniser un assuré sur ce seul fondement. En effet, le droit des assurances et les règles du droit civil commun ne permettront à l'assureur de se dégager de son obligation de garantie qu'en cas de fraude avérée. Or, l'on voit mal comment il pourrait se contenter pour cette démonstration du seul traitement automatisé du dossier de l'assuré[20].

Le directeur sinistre de Groupama évoquait un outil en phase test qui permettrait de mettre en évidence une anomalie. Et de conclure : « soit la manœuvre frauduleuse est *immédiatement révélée*[21], soit des investigations complémentaires d'enquêteurs certifiés sont nécessaires ». A en croire les termes de cette déclaration, le refus d'indemniser, serait susceptible de résulter en pratique d'un processus automatisé de traitement des données. S'il ne s'agit que d'un lapsus, il pourrait du moins révéler le souhait d'éviter à terme certains coûteux processus d'enquêtes. Les conditions dans lesquelles le RGPD admet le profilage permettent pourtant de douter de cette possibilité puisque dans ce cas, le traitement « humain » et la logique contradictoire devront être réintroduits sur demande de l'intéressé[22]. Il reste précisément intéressant de comprendre dans quelle mesure un traitement automatisé de données serait techniquement capable de justifier le refus d'indemniser, et d'alimenter éventuellement un dossier probatoire.

Question à l'actuaire :

Que pouvez-vous nous dire des analyses prédictives en matière fraude ? De quelles données ont-elles besoin ? Autrement dit, les modèles peuvent-ils être construits sur les seules données communiquées lors de la déclaration du risque et du sinistre, ou nécessitent-elles un apport complémentaire de données, et si oui lesquelles ? L'idée d'un algorithme établissant un dossier complet d'éléments prouvant la fraude dans un cas particulier est-elle réellement envisageable ? Autrement dit, pensez-vous qu'il soit possible de se passer d'une analyse in concreto de la situation de l'assuré, impliquant une intervention humaine d'enquêteurs ou d'analystes spécialisés ?

AC : On peut imaginer des données collectées en ligne. Par exemple voir sur Facebook qu'une personne est au ski, alors qu'elle est supposée être en arrêt maladie. Récemment, un assureur a refusé d'indemniser un assuré, en congé maladie pour neuf mois par suite de blessures aux cervicales, mais qui annonçait, quelques semaines après le début du congé, sa fierté d'avoir fini 7e à une course de 10 km (et sa prochaine participation à un semi-marathon) sur Twitter.

Les données GPS, utilisées pour l'instant dans un objectif de compréhension du risque, ne sont pas encore utilisées lors des accidents. Les marges d'erreur des boitiers sont importantes. Certains assureurs utilisent de l'analyse textuelle, en repérant des phrases ou des mots utilisés majoritairement par des fraudeurs. Il est aussi possible de croiser des données d'un grand nombre d'assurés pour repérer des garagistes qui surfacturent. Mais c'est toujours fait en complé-

[19] L'argus de l'assurance.com, Dossier spécial « Fraude », 4 / 8, Le big data sera-t-il la nouvelle arme antifraude ?, Eloïse Legoff - Publié le 03 septembre 2015 - http://bit.ly/2okhTwz, consulté le 22.11.17.

[20] L'AU n 39 en matière traitement de données ayant pour finalité la lutte contre la fraude, indiquait déjà que « les requêtes ou alertes détectées automatiquement doivent donner lieu à une analyse non automatisée par le personnel habilité de l'organisme [...] , le cas échéant des investigations complémentaires pourront être diligentées ».

[21] Surligné par nous.

[22] V. le droit de s'opposer au profilage en dehors du cas où il conditionne la conclusion du contrat, celui d'exiger une intervention humaine et d'avoir sur ce point un débat contradictoire (article 22).

ment, ou pour créer un score, un profilage, qui ensuite déclenchera ou pas une action (envoi d'un expert par exemple). L'an passé[23], François Nédey, directeur technique assurance de biens d'Allianz France, affirmait « Si nous soupçonnons de la connivence, nous allons *manuellement* regarder les données rendues publiques par l'utilisateur. Si nous confirmons un lien sur les réseaux sociaux, nous procédons alors à une enquête pour prouver la fraude ». Il s'agit de créer des scores, un outil d'aide à la décision, avec de l'analyse automatique de photos à la suite d'un sinistre, une extraction d'information dans un constat amiable, mais aussi de toutes sortes d'information, qui conditionneront l'envoi d'un expert ensuite.

DCS : On en conclura en conséquence qu'il n'y a pas finalement, dans les pratiques actuelles du moins, de « profilage » au sens du Règlement européen, c'est-à-dire de décision prise sur le seul fondement d'un traitement automatisé, dès lors que le scoring n'est qu'une alerte destinée à déclencher des investigations complémentaires. Il reste à savoir comment les assureurs gèreront la question de la conservation d'un assuré dans le fichier des fraudeurs potentiels ou confirmés. C'est en effet à eux, en vertu du principe d'accountability, que revient maintenant la charge d'apprécier le temps (strictement nécessaire) de la conservation de telles données[24].

Lutte contre la fraude et loyauté de la preuve.

RB : s'agissant de la fraude, avec les nouvelles technologies et les objets connectés en particulier, le double usage par l'assureur de la donnée personnelle (primo pour un suivi de la tarification, secundo pour la détection d'une fraude) semble facilité. Il en va de même avec le problème de la lutte contre le blanchiment d'argent sale.

L'Agence pour la lutte contre la fraude à l'assurance (Alfa) a pu indiquer qu'en 2015, les cas de fraude en assurances de dommages ont représenté un coût d'environ 2,5 milliards d'euros, soit 5% des primes, avec une prédominance en coût pour les dommages corporels car bien que représentant que 2% des cas, la fraude liée à ces dommages correspond à 47% de ces 2,5 milliards d'euros. Une source majoritaire de ces dommages corporels provient des accidents de la circulation.

En partageant entre assureurs les informations concernant les conducteurs et les véhicules sur la blockchain par exemple, une détection des tentatives de fraude multi-assurance pourrait être effectuée, allant même jusqu'au partage d'une note d'évaluation du conducteur qui le suivrait dans le temps pour que les assureurs puissent disposer d'un historique de son profil de conducteur, de son accidentologie ou de ses tentatives de fraudes.

D'un point de vue juridique, deux choses sont à concilier. D'un côté, la jurisprudence retient que la collecte de la preuve doit être loyale[25], à plus forte raison que l'assuré est présumé, en principe, de bonne foi par le Code des assurances. En outre, elle a condamné à plusieurs reprises des opérations de surveillance et de filature, jusqu'à l'intérieur du domicile, menées par les enquêteurs mandatés par l'assureur[26]. D'un autre côté, l'enquête privée devient de plus en plus intrusive et permanente, sans intervention humaine mais à l'aide d'objets connectés (boîtiers, caméras, capteurs, accouplés ou non à des smart contracts) où l'assureur et son personnel non qualifié se substituent directement à la profession d'enquêteurs privés. Or la loi n° 2003-239 du 18 mars 2003 pour la sécurité intérieure (JO 19 mars 2003) a rendu la formation obligatoire des professions d'enquêteurs privés, sauf pour ceux travaillant de façon interne dans une société (d'assurance) pour le seul compte de leur employeur. Le personnel de l'assureur n'est donc pas soumis aux obligations de qualification.

Questions à l'actuaire :

L'évolution rapide de certaines de ces technologies ne laisse-t-elle pas apparaître un risque d'aggravation d'un usage déloyal voire illicite de la preuve et/ou de l'enquête privée " digitalisée ", impliquant une collecte cachée de données personnelles à d'autres fins que le déclenchement de la garantie, par exemple pour la lutte contre le blanchiment et/ou contre la fraude à l'assurance ? L'actuaire peut-il ainsi tenter de limiter certaines pratiques ?

AC : De nombreuses technologies peuvent être intéressantes. Au Brésil, les services fiscaux utilisent des images satellites pour détecter la fraude de déclaration d'agriculteurs en extrapolant les quantités produites et donc les revenus. Un article récent[27] évoquait l'idée que les autorités fiscales pourraient utiliser les drones pour contrôler les propriétés. Mais cette utilisation de drones est vue « comme une ingérence dans la vie privée». Notons toutefois que certains assureurs commencent à utiliser Google Maps (et Street View) pour avoir des informations sur le logement d'une personne. Ces

[23] Dans un article de l'Est Républicain daté de septembre 2016 (http://bit.ly/2G4rE9x)

[24] Question que l'AU n39 de la CNIL réglait en proposant un mécanisme de conservation de la donnée « suspicion » en deux temps : 6 mois, d'une part, le temps de la qualifier (confirmer ou infirmer), puis 5 ans, d'autre part, une fois confirmée.

[25] Cass. ass. plén., 7 janv. 2011, nos 09-14.316 et 09-14.667, Bull. civ. ass. plén., n 1.

[26] Cass. 1re civ., 22 sept. 2016, n°15-24.015, Bull. civ. I ; adde Schulz R., Investigations portant atteinte à la vie privée : droit au respect de la vie privée et droit à un procès équitable, sous CEDH, 3 sect., 18 oct. 2016, n61838/10, RGDA 2016, n°12, p. 624 et s.

[27] « Le fisc interdit de drone pour contrôler les propriétés des contribuables » dans Le Figaro du 18 Janvier 2018 (http://bit.ly/2Dz19KZ).

méthodes relèvent toutefois de pratiques de gestionnaires de sinistres (pour vérifier les dégâts d'une tempête par exemple) ou de souscripteurs (pour voir la présence d'un garage) sur lesquels l'actuaire n'a pas la main. En effet, dans une compagnie d'assurance, l'actuaire cherchera à utiliser toutes les données accessibles, pour une mission dont il a la charge, comme proposer un nouveau tarif, proposer un indicateur de risque de fraude potentiel (de fraude à l'assurance). Une fois construit l'indicateur, c'est aux gestionnaires de sinistres de les utiliser : sur la base de suspicions (mauvais score donné par le modèle) le gestionnaire de sinistre pourra chercher des preuves d'une éventuelle fraude. Mais ça ne relève pas de la mission de l'actuaire.

Statistiques et recherche actuarielles

DCS : Pour finir sur cet aperçu du cycle de la donnée personnelle en assurance, venons-en au cœur même de l'activité actuarielle. Les informations personnelles d'un assuré ont en pratique vocation à alimenter les bases de données statistiques de l'actuaire chargé les modèles tarifaires d'une mutualité, voire de recherche actuarielles de dimension plus large, et ce souvent donc bien après le terme de la relation contractuelle qui a conduit l'assureur à les recueillir. Or, l'on entend ci et là, depuis l'annonce de l'entrée en vigueur prochaine du Règlement européen, que les compagnies d'assurance entendraient se défaire de leurs anciennes données plutôt que de risquer des sanctions pour non-conformité du traitement, ce qui risquerait d'affecter les bases de données actuarielles. Il semble opportun de clarifier un peu la situation sur ce point. La protection des données personnelles conduit à interroger leur exploitation en dehors du strict cadre de la relation d'assurance à plusieurs égards. Il s'agit d'abord d'élucider le fondement d'un tel traitement, de la question des garanties susceptibles d'être mises en place pour les personnes concernées, ensuite, et enfin de s'interroger sur la portée du « nouveau » droit à l'effacement (ou droit à oubli) consacré par le RGPD.

Le fondement du traitement ultérieur des données d'assurés. En principe, les données personnelles dont dispose l'assureur auront été collectées, soit au titre de la déclaration obligatoire des risques (C.ass., art. L. 113-2, 2°), soit, pour les données « sinistre », au titre de l'obligation de déclaration conditionnant le versement de la garantie (C. ass., art. L. 113-2, 4°). Il est admis que l'obligation légale vient sur ce point remédier à une asymétrie d'information incompatible avec la nécessité pour l'assureur d'évaluer le risque qu'on lui demande de prendre en charge et d'exécuter ses engagements. Si donc un droit lui est conféré sur les données d'un assuré, c'est dans le cadre du rapport bilatéral né du contrat qu'il passe avec celui-ci (risque individuel), ce qui ne préjuge en principe pas de ses prérogatives en matière de gestion de sa mutualité (risque collectif). L'analyse n'est pas vraiment différente si l'on se place du point de vue du Règlement. La collecte des données « risque » ou « sinistre » relèvera a priori d'un traitement « nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles ». De là à douter de la possibilité pour l'assureur de procéder ultérieurement à une exploitation actuarielle des données d'assurés, du moins sans consentement spécifique, il n'y a qu'un pas : parce que, d'une part, les prérogatives conférées à l'assureur par le Code des assurances ne se justifient que dans la mesure où il s'agit de gérer le lien contractuel avec l'assuré, et parce que, d'autre part, le Règlement soumet le traitement des données au principe de limitation des finalités (art. 5.1, b).

L'assureur devrait néanmoins pouvoir invoquer son intérêt légitime (à équilibrer économiquement sa mutualité) pour fonder une exploitation des données au-delà du contrat (RGPD, art. 6.1, f). Les intérêts et droits fondamentaux des assurés pourraient prévaloir sur l'intérêt de l'assureur s'ils « ne s'attendent raisonnablement pas à ce traitement ultérieur »[28]. Cela étant, ils devront en tout état de cause être informés de ce traitement[29] et notamment du fait qu'ils disposent dans cette hypothèse (fondement sur l'intérêt légitime) d'un droit d'opposition (art. 21).

Plus directement, le Règlement autorise un traitement à des fins différentes du moment que celles-ci sont compatibles avec les finalités initiales (art. 5.1, b). Cette compatibilité s'apprécie au regard de plusieurs critères[30], dont le premier cité est le lien entre les finalités successives, ce qui s'agissant d'apprécier un risque individuel et le risque collectif de la mutualité dans lequel il s'est inséré ne semble pas faire de doute. Ce lien sera cependant insuffisant à lui seul pour justifier la compatibilité, et devra s'accompagner d'autres circonstances, notamment l'existence de garanties appropriées, « dont le chiffrement ou la pseudonymisation ».

Par ailleurs, le Règlement reprend la solution de la Directive de 1995[31] selon laquelle le traitement ultérieur à des fins statistiques, historiques ou scientifiques est réputé compatible avec les finalités initiales du traitement (art. 5.1, b, in fine). Pour l'instant, le droit français n'envisage à ce titre que les études et bases de données d'intérêt public. Or, si l'on en croit le considérant 159 dudit Règlement[32], il n'est pas totalement exclu que cette hypothèse puisse recouvrir les études

[28] Considérant 47

[29] Enoncés par l'article 6.4 qui en dénombre cinq : lien entre les finalités initiales et ultérieures, contexte de la collecte, nature des données, conséquences pour l'intéressé et enfin garanties appropriées.

[30] Enoncés par l'article 6.4 qui en dénombre cinq : lien entre les finalités initiales et ultérieures, contexte de la collecte, nature des données, conséquences pour l'intéressé et enfin garanties appropriées.

[31] Article 6.

[32] Il y est notamment dit que « Aux fins du présent règlement, le traitement de données à caractère personnel à des fins de recherche

et statistiques actuarielles réalisées pour le compte des assureurs puisque, nous dit-on, le terme de recherche « devrait être interprété au sens large et couvrir, par exemple [...] la recherche appliquée et celle financée par le secteur privé ». La question d'une telle assimilation devra être discutée, mais si elle était admise, le travail actuariel resterait en tout état de cause soumis à l'exigence de mesures techniques et organisationnelles garantissant les droits et libertés de la personne concernée. C'est ce que rappelle l'article 89.1 in fine, tout en ajoutant que « chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière ».

L'existence de garanties appropriées, dont la pseudonymisation. En définitive, pour pouvoir être exploitée ultérieurement, la donnée d'assuré aurait donc au premier chef vocation à être « modifiée » pour perdre en tout ou partie son caractère identifiant (anonymisation ou pseudonymisation).

S'agissant de la donnée anonyme, celle-ci est par hypothèse exclue du champ de la protection des données dites « personnelles ». Ce tout simplement parce que les deux termes sont antinomiques, la donnée personnelle supposant la faculté d'identification. En anonymisant les données d'assuré, l'assureur devrait pouvoir les soumettre sans difficulté au travail actuariel. Cependant, compte tenu de ce qui a pu être démontré sur le pouvoir du croisement d'informations, l'anonymisation de données « risque » ou « sinistre » parait difficilement praticable. La technique de la « pseudonymisation », quant à elle, n'enlève pas son caractère personnel à la donnée, mais permet d'en alléger le régime. Le terme désigne une sorte d'anonymisation réalisée a minima, en ce qu'elle n'est pas définitive. Le Règlement définit la pseudonymisation comme le « traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable » (art. 4.5).

Il convient néanmoins d'observer que, dans le cas particulier des études ou bases de données d'intérêt public, il est précisé que la contrainte liée à l'identification des personnes concernées n'existe que « dans la mesure où ces finalités peuvent être atteintes de cette manière ». Si donc le traitement actuariel bénéficiait de cette qualification, il pourrait être dispensé de l'anonymisation ou de même de pseudonymisation chaque fois que ladite technique viendrait à l'entraver.

Mais en définitive, que la nécessité de faire perdre aux données leur caractère identifiant soit absolue ou relative, cela soulève la question de la compatibilité de cette exigence avec l'exploitation actuarielle des données d'assurés.

Question à l'actuaire :

Les méthodes de travail de l'actuariat d'assurance sont-elles ou peuvent-elles être construites à partir de données anonymisées ? Ou de données simplement pseudonymisées ? Sinon, quels sont les avantages de disposer de données permettant l'identification directe ou indirecte de la personne de l'assuré ?

AC : Encore une fois, le but n'est jamais d'identifier les gens. Latanya Sweeney a étudié[33] un exemple resté fameux (et relatif à des problèmes d'assurance), à savoir les déclarations annuelles obligatoires de Group Insurance Commission (GIC) au Massachusetts, aux États-Unis. Cette institution semi-publique a pour mandat d'offrir des contrats d'assurance santé aux employés de l'État du Massachusetts (135 000 personnes), mais avait obligation de fournir des statistiques agrégées sur son portefeuille. Par date de naissance, genre et code postal, l'assureur devait fournir des informations sur les visites et les remboursements médicaux. En utilisant les données électorales de la ville de Cambridge, six personnes partageaient la date de naissance du Gouverneur William Weld, trois seulement étaient des hommes, et il a été aisé de retrouver son code postal. Instantanément, il a été possible d'identifier sans erreur aucune le Gouverneur, et d'avoir des informations non publiques sur son état de santé.

Pseudonymiser les données est un exercice complexe. En particulier en assurance habitation où il est nécessaire de connaître le lieu de résidence précis. En assurance automobile, à partir de l'âge, du code postal et du modèle de véhicule, plus de la moitié du portefeuille serait identifié avec certitude dans une ville de taille raisonnable. Il n'y a aucun avantage, aucun intérêt à le faire. Mais le fait est que les actuaires peuvent très souvent identifier les données sur lesquelles ils travaillent. Des travaux théoriques sont en cours sur l'utilisation de techniques d'encryptages pour la construction de modèles prédictifs. Mais l'exercice est complexe (et trop technique) pour la discussion que nous avons ici.

---

scientifique devrait être interprété au sens large et couvrir, par exemple, le développement et la démonstration de technologies, la recherche fondamentale, la recherche appliquée et la recherche financée par le secteur privé »

[33] En France, le droit n'est en effet que « formellement » nouveau dès lors que la possibilité de demander la suppression de ses données existe « substantiellement » depuis la Loi Informatique et Libertés de 1978 : il s'exerce alors dans le cadre de la mise en œuvre d'un droit plus large incluent le droit de rectification (LIL, art. 40). V. également, le droit au déréférencement consacré dans l'arrêt Google Spain rendu par la CJUE le 13 mai 2014 (affaire C-131/12) sur le fondement de la Directive de 1995.

Le droit à l'effacement.

DCS : Il reste en tout état de cause une autre difficulté. En effet, protéger les données en les encryptant ou en les dépouillant de tout ou partie de leur caractère identifiant suppose de les avoir conservées, du moins d'avoir été en mesure de le faire. Or, la question du « nouveau »[34] droit à l'effacement consacré par le Règlement européen est de nature à affecter les bases de données d'assurance s'il devait être exercé massivement dans une mutualité. Comme Pascal l'a montré il y a déjà longtemps avec la loi des grands nombres, la fiabilité de la prédiction suppose de pouvoir se fonder sur un nombre élevé de données d'expérience. Les conditions d'exercice du droit à l'effacement (RGPD, art. 17) permettent cependant de nuancer l'observation.

D'abord, parce que le droit est exclu si l'on trouve face à un traitement relevant de fins statistiques ou de recherche scientifique (RGPD, art. 89) dont l'effacement « est susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs ». La question de la qualification de l'actuariat d'assurance ressurgit ici. Si l'on admet que cette activité est couverte par l'article 89, l'assureur n'a pas à redouter le droit à l'effacement. Le droit ne pourrait jouer que si cela ne constitue pas un véritable obstacle à ses études actuarielles. Il restera peut-être à s'entendre sur l'appréciation de la gravité de l'entrave...

Ensuite même si le droit à l'effacement n'était pas exclu (actuariat ne relevant pas de l'article 89), la menace que ce droit constitue doit être relativisée parce qu'il suppose en substance que le traitement ne soit pas ou plus fondé. C'est le cas lorsque ses finalités sont épuisées, ou lorsque l'intéressé a exercé son droit d'opposition sans que le responsable du traitement ait de motifs impérieux de s'y opposer. Or, l'exploitation actuarielle des données d'anciens assurés est n'est pas a priori une finalité qui a vocation à s'épuiser rapidement. Il faudrait pouvoir imaginer un risque que l'on aurait plus de raison légitime d'observer, parce qu'il aurait disparu, ou parce que la mutualité aurait radicalement changé de composition... En bref, l'assuré qui voudrait exercer son droit à l'effacement devrait donc justifier au préalable d'un droit d'opposition à l'exploitation ultérieure de ses données (art. 21).

La vigueur du droit d'opposition de l'assuré dépendra alors du fondement que l'on reconnaitra au traitement actuariel. De deux choses l'une. Soit l'on retient l'« intérêt légitime » de l'assureur à alimenter les statistiques de sa mutualité, et alors le droit d'opposition suppose non seulement d'invoquer des « raisons tenant à sa situation particulière », raisons qui devront encore être confrontées aux « motifs légitimes et impérieux » de l'assureur (art. 21.1). Soit l'on considère que l'actuariat d'assurance relève de l'article 89, et le droit d'opposition supposera toujours que l'assuré puisse justifier de « raisons tenant à sa situation particulière », sous réserve cette fois que le traitement ne soit pas « nécessaire à l'exécution d'une mission d'intérêt public ». Or, si l'on admet l'application de l'article 89 aux études actuarielles et aux statistiques assurantielles, c'est qu'on leur aura déjà en principe reconnu un certain intérêt public.

On le voit la marge de manœuvre en matière de traitement post-contractuel des données d'assurés suppose de trancher deux questions. L'une est de déterminer si le fait d'accorder à l'assuré un droit à l'effacement de ses données après le terme du contrat (et l'écoulement des délais de prescription) compromettrait « gravement » les études actuarielles. L'autre est de se demander dans quelle mesure il est légitime d'assimiler ces études à la recherche scientifique, voire à des missions d'intérêt public, et partant de leur accorder un statut dérogatoire au regard des droits des personnes concernées[35].

Question à l'actuaire :

Quel regard portez-vous sur ce droit à l'effacement (ou à l'oubli) et la menace qu'il représente pour l'actuariat d'assurance ? Pensez-vous qu'il serait légitime d'assimiler le travail que l'actuaire effectue sur les données d'assurés à la notion de « recherche scientifique » ou de « fins statistiques » visée par le RGPD, afin de limiter les droits des assurés sur leurs données ? Dans quelle mesure peut-on considérer que les études utiles aux assureurs, personnes privées, sont assimilables à celles faites dans l'intérêt public ?

AC : Pour illustrer le débat, on peut regarder les pratiques de certains assureurs. La MAIF par exemple propose une « charte numérique »[36], avec un premier volet sur la « protection de données personnelles ». On y parle de « respecter », d'« être transparent », de « sécuriser » et surtout d'« oublier ». Pour être plus précis, le dernier point est énoncé

[34] En France, le droit n'est en effet que « formellement » nouveau dès lors que la possibilité de demander la suppression de ses données existe « substantiellement » depuis la Loi Informatique et Libertés de 1978 : il s'exerce alors dans le cadre de la mise en œuvre d'un droit plus large incluent le droit de rectification (LIL, art. 40). V. également, le droit au déréférencement consacré dans l'arrêt Google Spain rendu par la CJUE le 13 mai 2014 (affaire C-131/12) sur le fondement de la Directive de 1995.

[35] A l'heure où ce « dialogue » est en voie de publication, le projet de loi français d'adaptation de la loi de 1978 au RGPD (AN, n°490, 15e législature) ne semble pas aller dans cette voie. L'article 12 du projet prévoit en effet que « les conditions et garanties appropriées prévues à l'article 89 du règlement (UE) 2016/679 sont déterminées *par le code du patrimoine et les autres dispositions législatives et réglementaires applicables aux archives publiques*. Elles sont également assurées par le respect des normes conformes à l'état de l'art en matière d'archivage électronique », ce qui indique une vision plus restrictive des hypothèses de l'article 89 qui se limiteraient plutôt aux activités des personnes ou services publics.

[36] Charte Numérique de la MAIF, en ligne le 2 février 2018 (http://bit.ly/2AhVOpv)

de la manière suivante « Dans une société de la mémoire, le droit à l'oubli devient un droit fondamental. Chacun peut nous demander à tout instant la suppression des données qui le concernent, dans le respect de nos obligations de conservation ». Cette déclaration peut sembler généreuse, dans l'esprit des directives récentes, mais si on pousse le raisonnement jusqu'au bout, que faire si tout le monde utilise ce droit ? Comment les actuaires vont-ils pouvoir tarifer si tout le monde exerce ce droit, et qu'il n'existe plus de données pour faire des calculs statistiques. Il est toutefois mentionné une « obligation de conservation ». Il existe effectivement quelques obligations : dans le code des assurances, des délais de prescriptions sont mentionnés. Par exemple, l'article L.114-1 dit que « toutes actions dérivant d'un contrat d'assurance sont prescrites par deux ans à compter de l'événement qui y donne naissance ». Autrement dit, il y a une obligation de conserver pendant deux ans. Ce délai de deux ans fait l'objet de deux exceptions : « la prescription est portée à dix ans dans les contrats d'assurance sur la vie lorsque le bénéficiaire est une personne distincte du souscripteur et, dans les contrats d'assurance contre les accidents atteignant les personnes, lorsque les bénéficiaires sont les ayants droit de l'assuré décédé » (alinéa 4) et « pour les contrats d'assurance sur la vie… les actions du bénéficiaire sont prescrites au plus tard trente ans à compter du décès de l'assuré » (alinéa 5). Il existe donc en effet des obligations pour conserver des données.

Mais au-delà de cette « obligation » de conservation, l'article 89 du RGPD mentionne surtout un « droit de conservation » correspondant à un « traitement à des fins archivistiques », en lien avec un « intérêt public ». Ce point est important et essentiel pour les actuaires, mais pas seulement. Certains mécanismes d'assurance ont été mis en place dans un « intérêt public », comme l'assurance contre les catastrophes naturelles (loi du 13 juillet 1983). Pour améliorer le système de couverture et d'indemnisation, il pourrait être dans l'intérêt public de faire des études sur certaines catastrophes récentes. Or pour étudier ces risques rares, il convient d'utiliser des sinistres sur une longue période temporelle. Pour comprendre le risque centenaire, utiliser deux ans d'historique n'aidera pas beaucoup. Or combien d'assureurs ont encore des données personnelles relatives aux tempêtes de 1999 (vieilles de moins de 20 ans) ? En 2005[37], la CNIL avait introduit le concept de « données définitives », (présentant un intérêt historique, scientifique ou statistique justifiant qu'elles ne fassent l'objet d'aucune destruction). Ne faudrait-il pas imaginer un système permettant d'archiver, au niveau national, certaines de ces données, à des fins de recherche ? S'il existe une distinction entre les données détenues par des personnes « privées » et des personnes « publiques », comme traiter le cas des catastrophes naturelles, qui fonctionne sur un mécanisme d'acteurs privés, mais dont la prime est fixée par décret gouvernemental ?

[37] Délibération n°2005-213 du 11 octobre 2005