



HAL
open science

La data dans l'assurance santé et prévoyance

Tony Sagaliapidine

► **To cite this version:**

Tony Sagaliapidine. La data dans l'assurance santé et prévoyance. domain_shs.info.docu. 2020. mem_03172166

HAL Id: mem_03172166

https://memic.ccsd.cnrs.fr/mem_03172166v1

Submitted on 17 Mar 2021

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



le **cnam**
intd

LA DATA DANS L'ASSURANCE SANTE ET PREVOYANCE

Mémoire pour l'obtention du
Master Sciences humaines et sociales
mention humanités numériques
Parcours Mégadonnées et analyse sociale
(MEDAS)

Tony SAGALIAPIDINE

Date et lieu de la soutenance

- 09 Septembre 2020
- CFA IDFSAINTE-DENIS (93)

Membres du jury

- CHARTRON Ghislaine, présidente du jury
- ARRUABARENA Béatrice, présidente du jury
- KILANI Karim, directeur de mémoire
- CHARTRAIN Pauline, tutrice entreprise

Promotion (2018-2020)



Paternité Pas d'Utilisation Commerciale - Pas de Modification

« Ce métier est une quête infinie, qui nous rend parfois « schizophrène » à force de faire et de défaire tout ce qu'on met en place, est aussi la plus intéressante. »

Sarah Alem

Le but de ce mémoire est de mettre en exergue l'impact de l'utilisation de la donnée dans le secteur de l'assurance santé et prévoyance.

Dans un premier temps nous ferons l'état des lieux des différentes connaissances et des pratiques existantes autour de l'utilisation de la donnée. Puis, à travers mon expérience au sein de l'assureur GENERALI, nous verrons comment mettre en place un outil d'aide à la décision. Cet outil permettra de suivre la rentabilité de produit d'assurance vendu par délégation.

Mots clés : assurance, Big Data, reporting, suivi activité

The purpose of this thesis is to highlight the impact of the use of data in the health insurance and provident sector.

First, we will take stock of the different knowledge and existing practices around the use of data. Then, through my experience with the insurer GENERALI, we will see how to set up a decision support tool. This tool will make it possible to monitor the profitability of insurance products sold by delegation.

Keywords: activity monitoring, Big Data, insurance, reporting

REMERCIEMENTS

Je tiens tout d'abord à adresser mes remerciements à tous ceux qui m'ont permis de réaliser mon apprentissage au sein de l'assureur Generali France.

J'exprime ma profonde gratitude à Madame CHARTRAIN Pauline, manager, et Madame KANZARI Nourimane, responsable d'études actuarielles, de m'avoir accueilli au sein de leur service, de m'avoir fait confiance et je les remercie grandement pour leurs disponibilités et toutes les connaissances qu'elles ont pu m'apporter au cours de ces deux années passées.

J'adresse également mes remerciements à l'ensemble du département de la Prévoyance & Santé pour son professionnalisme, sa disponibilité et sa convivialité.

Je remercie M. KILANI, tuteur pédagogique, pour sa disponibilité et ses précieux conseils.

Enfin, je tiens à adresser ma reconnaissance au Conservatoire National des Arts et Métiers (CNAM), ainsi qu'à l'ensemble du corps professoral pour les nombreuses connaissances que j'ai pu acquérir lors de ma scolarité.

Table des matières

Remerciements	5
Table des matières	6
Introduction	8
Partie I - Etat de l'art	9
I. <i>Cerner les fondamentaux et les enjeux de la data</i>	9
II. <i>La place de la donnée au cœur de ce paradigme</i>	11
a. La donnée comme point de départ	11
b. Le cycle de vie de la donnée	11
c. La qualité de la donnée	18
d. La marchandisation des données	18
e. Les sources de données se diversifient	19
III. <i>Identifier l'écosystème du Big data</i>	20
a. Architectures et solutions Data	20
b. Les algorithmes	23
IV. <i>L'impact du Big data sur le secteur de l'assurance</i>	24
a. L'environnement spécifique de l'assurance	24
b. Les solutions de Big Data : Quels impacts en assurance ?	25
c. Des impacts métiers pour l'assureur	27
V. <i>Passer de la théorie à la pratique</i>	29
a. Comment concevoir un projet data ?	29
b. Les méthodes Agile	30
c. Les méthodes traditionnelles	30
d. L'utilisation des méthodes Agile pour les projets Big data	31
VI. <i>Cerner les enjeux réglementaires et juridiques</i>	32
a. Le Règlement Général sur la Protection des Données	32
b. Une protection des données règlementée par la CNIL	32
Partie II - GENERALI Assurance	34
I. <i>Historique</i>	34
II. <i>Generali France</i>	34
III. <i>Présentation du service</i>	35
L'assurance Santé & Prévoyance	36
I. <i>Définition de l'Assurance Santé</i>	36
a. Comment fonctionnent les remboursements en santé ?	37
b. Plusieurs niveaux de couverture	38
II. <i>Définition de l'Assurance Prévoyance</i>	40
a. Les principaux types de contrats de prévoyance	40
b. Prévoyance individuelle ou collective	41
La délégation de gestion	43
I. <i>La délégation à l'opportunité pour les assureurs</i>	43

II. Mise en place d'applications de gestion des délégataires	43
a. Création d'applications de suivi de gestion des délégataires	43
b. Assurer la conformité des activités déléguées	43
III. Les délégataires chez Generali	44
Partie III - Solution analytics - Mise en place des suivis	45
I. Suivi de rentabilité.....	45
II. Analyse du portefeuille.....	48
III. Analyse de la consommation	49
IV. Automatisation des Solutions analytics	50
Conclusion.....	52
Table des figures	53
ANNEXE I.....	54
ANNEXE II.....	55
Bibliographie	56

INTRODUCTION

Dans le cadre du Master MégaDonnées et Analyse Sociale du CNAM, j'ai réalisé mon année en apprentissage au sein de l'assureur GENERALI.

Tout en me formant au métier de *Data Analyst*, les missions que j'ai pu réaliser m'ont permis de :

- Découvrir les différentes étapes de l'exploitation et de la gestion d'une base de données (administration, automatisation, développement des procédures etc.) ;
- Voir comment se déroule une veille technologique sur des Systèmes de Gestions de Base de Données (SGBD) ;
- D'assister au dimensionnement des bases.

Ainsi, j'ai pu participer à la mise en place d'un suivi d'activité et relever les enjeux que celle-ci comporte vis-à-vis de la politique menée par l'entreprise.

Aujourd'hui, les compagnies d'assurance gèrent de plus en plus de grandes quantités de données. Elles sont issues principalement des contrats d'engagement mutuel entre l'assureur et l'assuré. Jusqu'à présent, un nombre relativement limité de données externes étaient nécessaires pour que l'assureur opère son métier de gestionnaire des risques. Des statistiques, des données de place, des informations économiques, des éléments sur les biens ou les personnes assurées permettent désormais de beaucoup mieux appréhender et évaluer le risque. En cela, le Big Data représente une opportunité pour les assureurs d'introduire de nouveaux paramètres dans la connaissance, l'évaluation et la gestion du risque.

Ainsi, nous verrons au travers de ce rapport la problématique de l'utilisation du Big Data au service de l'assurance. Puis, nous présenterons les missions réalisées au cours de cette année, avant de faire un bilan.

PARTIE I - ETAT DE L'ART

I. CERNER LES FONDAMENTAUX ET LES ENJEUX DE LA DATA

Par définition, l'état de l'art fait acte de recueil de connaissances en montrant explicitement une synthèse diachronique de la dynamique du processus avec citation dynamique de tous les éléments concernés et ce, afin de démontrer la créativité conceptuelle et opérationnelle projetée dans un mémoire. Nous devons considérer deux perspectives au cours de cet état de l'art : la description des composantes et la saisie de la dynamique du processus qui a présidé au mode évolutif des concepts et principes concernés.

Il convient, tout d'abord, de situer le contexte dans lequel nous nous trouvons : celui de l'assurance Santé & Prévoyance. Tout au long de cet état de l'art seront définis différents termes liés à la Data. Ensuite, nous étudierons le développement de la pratique de la data science dans une compagnie d'assurance. Puis, nous développerons les cas d'usages spécifiques au monde de l'assurance. Après avoir posé ces quelques jalons, nous pourrions prioriser et proposer des solutions pour notre cas d'usage à GENERALI Assurance.

LA DATA DANS L'ASSURANCE SANTE ET PREVOYANCE : Le reporting, un outil d'aide à la décision

Est un sujet pluraliste. En effet, nous pouvons traiter ce sujet d'un point de vue stratégique, faisant appel au marketing sectoriel, etc. mais nous choisissons l'approche du point de vue technique, en exposant le bénéfice de l'utilisation de la data dans le secteur de l'assurance.

Comme annoncé précédemment, il conviendrait tout d'abord de définir en premier lieu quelques termes nécessaires à la compréhension de notre sujet. En effet, lorsque l'on effectue des recherches liées à la data, celles-ci aboutissent tous sur un concept clé : **le Big Data**.

Il existe une multitude de définitions pour caractériser le Big data, nous avons donc choisi l'une d'entre elles qui fait référence, celle la plus largement utilisée dans l'industrie :

« Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation. » (GARTNER, 2012)

Nous pouvons retenir de cette définition que le Big Data est constitué de 3 composantes : Volume, Vitesse et Variété. Autrement appelé : les « 3V du Big Data ». La Data Scientist H. ISSARANE (2019) définit ces composantes de la manière suivante :

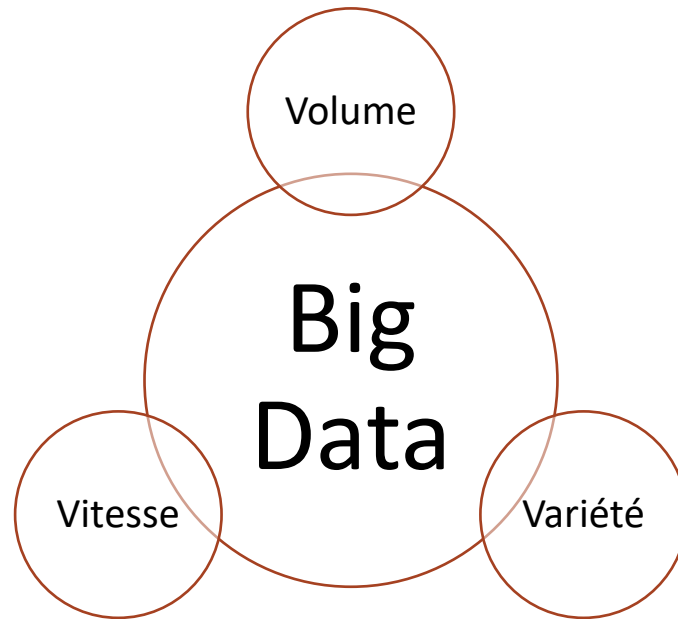


Figure 1 - Les 3V du Big Data

- **Volume** : Le volume décrit la quantité de données utilisées et générées par les entités ou des personnes. Aujourd’hui, les entreprises, tous secteurs d’activité confondus, sont confrontés à une croissance exponentielle du stockage de données.
- **Vitesse** : Les entreprises ont besoin aujourd’hui d’une grande vitesse de traitement associée à une évolution très rapide des données disponibles.
- **Variété** : Des données non structurées de tout type (données numériques, textes, images, vidéos, son), sont aussi utilisées par les entités.

Nous pouvons retenir que le Big Data n'implique pas seulement une grande volumétrie et typologie de données, il s’agit plutôt d’un schéma permettant d’améliorer la compréhension et l’analyse de celles-ci. Nous pouvons même traduire ce schéma par une 4^{ème} Valeur du Big Data : **la valeur** potentielle attendue de leur exploitation.

II. LA PLACE DE LA DONNÉE AU CŒUR DE CE PARADIGME

a. La donnée comme point de départ

Étymologiquement, le mot data est dérivé du latin *dare*, qui signifie « donner ». En ce sens, les données sont des éléments bruts qui peuvent être extraites des phénomènes - mesurés et enregistrés de diverses manières (STROWEL, 2018). Autrement dit, la donnée peut être de nature bien différente : texte, chiffre, son, image ou vidéo. Elle constitue une information élémentaire codée sous un format numérique et employée pour décrire un objet, une personne ou encore un fait.

Dans les entreprises où le digital prend une importance considérable, la donnée est l'élément de base, essentielle à tous les systèmes et processus opérationnels en constituant la matière première de leurs activités.

Dès lors, les données représentent une véritable valeur économique et chacun prend conscience des différentes utilisations possibles qu'il peut en faire. La maîtrise des données devient alors un enjeu stratégique pour les entreprises.

Pour les compagnies d'assurances, l'accès aux informations est primordial : l'exploitation des données offre ainsi une meilleure connaissance des clients et de leurs comportements. Cette connaissance permet par exemple de proposer aux clients des tarifs mieux adaptés et plus justes, ou encore proposer de nouveaux services de prévention.

b. Le cycle de vie de la donnée

« Une donnée vit : elle naît, grandit, se reproduit, voyage, et prend sa retraite. »

Jean Laurent GRANIER, Président-Directeur Général de GENERALI France

Les données constituent une vraie richesse, pour peu qu'elles soient collectées et convenablement stockées dans des systèmes adéquats. Ceux-ci doivent être en capacité de les mettre à disposition à des spécialistes qui sauront les utiliser à bon escient. Ainsi, les données transitent au sein d'une entreprise par plusieurs étapes. Nous utiliserons le cas de la compagnie d'assurance GENERALI pour illustrer ce cycle de sept étapes illustrées ci-dessous.



Figure 2 - Cycle de vie de la donnée chez GENERALI

Chacune de ces étapes est importante, et si elles sont réalisées de façon conforme, les données prennent alors de la valeur et deviennent une réelle source de business pour les compagnies d'assurances.

- **Etape 1 : La collecte**

Une donnée naît lorsqu'elle est collectée ou saisie. Et pour naître dans de bonnes conditions, on y applique des méthodes et des obligations.

Les données peuvent être donc collectées directement auprès des personnes concernées. La collecte s'effectue à l'aide d'un formulaire sous format papier ou électronique (sur une application et/ou site internet par exemple). Grâce à ce formulaire, la collecte des données est dite structurée.

Il se peut aussi que les données soient non-structurées. C'est le cas des données générées par les mails, les présentations PowerPoint, les documents Word/Excel/PDF, ou encore les logiciels de collaboration ou de messagerie instantanée. Ce sont également des données générées sous des formats images JPEG, fichiers audio MP3, ou encore des fichiers vidéo.

Les données peuvent également être récupérées auprès de sources diverses. Par exemple, lors d'une transmission de données entre différents services, différents métiers ou encore auprès d'organismes extérieurs possédant des banques de données (open data).

Les obligations lors du processus de collecte

Le processus de collecte des données à caractère personnel est strictement encadré par la loi Informatique et Libertés et le règlement général sur la protection des données (RGPD) énonce les obligations auxquelles doivent répondre les professionnels.

Finalité : informer sur les finalités du traitement des données

Les formulaires de collecte de données ainsi que les contrats doivent toujours comporter une clause qui explique à la personne concernée quelles sont les finalités poursuivies par le traitement, qui est le responsable de traitement, si les données seront transmises à des sous-traitants, la durée de conservation de leurs données et des droits dont ils disposent pour accéder à ces dernières (droit d'accès, droit de rectification, droit de portabilité, droit d'opposition et droit de suppression).

Pertinence : collecter ce qui est strictement nécessaire

Seules les données strictement nécessaires à la réalisation de l'objectif peuvent être collectées : c'est le principe de minimisation de la collecte. Le responsable de traitement ne doit donc pas collecter plus de données que ce dont il a vraiment besoin. Il doit également faire attention au caractère sensible de certaines données.

Consentement : recueillir le consentement pour chacune des finalités

Pour être démarchée, la personne doit avoir acceptée le principe du démarchage et pour une finalité précise. Il faut donc veiller à avoir recueilli l'accord préalable de la personne concernée. Il est indispensable d'obtenir son consentement pour chacune des finalités du traitement des données.

Sécurité : sécuriser les processus de collecte des données

Le processus de collecte des données doit rester confidentiel et l'accès physique aux données et aux traitements doit être réglementé, d'autant plus s'il s'agit de données de santé. Les outils informatiques doivent également être sécurisés par des mots de passe et les données cryptées tout en restant accessibles et traçables par le responsable de traitement.

- **Etape 2 : Le contrôle**

La qualité des données passe par le contrôle de ces dernières. Cette seconde étape impacte le reste du parcours de vie de la donnée. Il est donc essentiel qu'elle se déroule de façon conforme.

Il est important de contrôler l'exactitude des données récoltées. En effet, une donnée exacte est une donnée de qualité et la qualité des données est fondamentale pour servir les enjeux de l'Entreprise.

L'accès aux informations devient primordial pour les assureurs qui, grâce à l'exploitation des données, ont dès lors une meilleure connaissance des clients et de leurs comportements. Cette connaissance nous permet par exemple de proposer à nos clients des tarifs mieux adaptés et plus justes et de les indemniser dans les meilleures conditions.

Les obligations lors du processus de contrôle

Le contrôle des données à caractère personnel est strictement encadré par la loi Informatique et Libertés et le règlement général sur la protection des données (RGPD) énonce les obligations auxquelles doivent répondre les professionnels.

Exactitude : contrôler l'exactitude des données collectées

A cette deuxième étape de la vie de la donnée, il est nécessaire que le responsable du traitement contrôle que les données soient exactes, complètes, et si nécessaire, qu'elles soient mises à jour.

Sécurité : gérer les failles de sécurité liées à des données

En cas de violation de données à caractère personnel, le responsable de leur traitement doit le notifier à l'autorité de protection des données dans les 72 heures. Les personnes concernées doivent également en être informées si cette violation est susceptible d'engendrer un risque élevé pour leurs droits et libertés.

- **Etape 3 : Le traitement**

La phase de consommation et de traitement est l'étape la plus intéressante du parcours de la donnée : C'est à ce moment qu'elle sera utilisée pour délivrer toutes ses informations à la bonne personne, au bon moment et pour une action précise.

Le traitement des données représente toutes les opérations effectuées sur des données à caractère personnel, que le procédé soit automatisé ou non.

Les obligations lors du processus de traitement

Responsable : définir un responsable de traitement des données

Il est indispensable de définir clairement qui fait quoi, qui en est responsable, veiller à ce que les partenaires respectent les mêmes règles de conduite et soient en conformité avec la réglementation.

Droits : gérer le droit d'accès des personnes aux données

Les personnes qui voient leurs données personnelles récoltées disposent de droits afin de garder la maîtrise sur ces dernières. Le responsable de traitement doit pouvoir, à la première demande :

- ✓ restituer les données collectées (droit de portabilité),
- ✓ les rectifier (droit de rectification),
- ✓ les supprimer (droit de suppression ou droit à l'oubli numérique),
- ✓ stopper les traitements (droit d'opposition),
- ✓ porter les données auprès d'une personne désignée par le demandeur dans la limite des obligations et durées de conservation légale.

Registre : mettre à jour le registre des activités de traitement

Le traitement de gros volume de données ou encore de données de santé implique de tenir à jour un registre qui décrit les traitements effectués. Ce registre des traitements devra être fourni à la CNIL qui répertorie toutes les données collectées.

Sécurité : sécuriser les processus de consommation et de traitement des données

La consommation et le traitement des données doivent rester confidentiels et l'accès physique aux données et aux traitements doit être réglementé, d'autant plus s'il s'agit de données de santé.

Les outils informatiques doivent également être sécurisés par des mots de passe et les données cryptées, tout en restant accessibles et traçables par le responsable de traitement.

• **Etape 4 : L'évolution**

Comme les humains, les données connaissent également des périodes de transformation. Elles évoluent au fil de la vie du client qu'elle permet d'identifier ou encore pour rester en conformité avec la législation.

Une donnée n'est pas figée dans le temps. Elle peut être enrichie, supprimée d'une base de données, déplacée, voir sa structure se modifier... Les droits que peuvent exercer les personnes sur leurs données personnelles impactent directement l'évolution de ces données.

L'évolution de la donnée, de sa première collecte à ce qu'elle est devenue, est particulièrement intéressante. Cette évolution est utilisée et traduite en statistiques dans le but de tirer des constats et de mettre en place des actions.

Les obligations lors de l'évolution des données

Veille : s'adapter à l'évolution de la législation

Il faut être vigilant quant à la réglementation qui évolue sans cesse vers une responsabilisation de plus en plus importante des acteurs qui traitent de la donnée. Le renforcement des droits aux personnes implique d'être agile et de se tenir prêt à apporter des modifications aux données que l'on possède.

Changement : suivre l'évolution des données

Lorsque la situation ou une information relative à un individu, dont les données ont été récoltées, changent, elles doivent alors être remises à jour.

Exemple : un changement de nom suite à un mariage ou un divorce, un décès, etc.

- **Etape 5 : Le transport**

Le transport représente les déplacements que sera amenée à vivre la donnée. Elle pourra connaître différents lieux de stockage, différents propriétaires et même différents moyens de transport. Et parce qu'une donnée est précieuse, des règles de déplacement sont définies.

L'échange de données ne semble pas connaître de frontière, et pourtant l'impact de cette mobilité sur l'environnement n'est pas virtuel.

Au quotidien, nous effectuons des recherches sur Internet ou encore échangeons des mails avec ou sans pièce jointe.

Ces échanges dématérialisés, qui donnent le sentiment de ne pas avoir d'impact, ont en réalité des conséquences sur l'environnement.

La donnée est transportée et mise à disposition de plusieurs personnes. Ce partage est physique, il est lourd et engendre des coûts pour l'entreprise.

Les bonnes pratiques de transport de données

Pour limiter les impacts écologiques de nos mails, des gestes simples peuvent être mis en place : Outre les conséquences sur l'environnement, le transport des données doit être sécurisé selon le niveau de classification de la donnée concernée.

Quatre niveaux de classification sont identifiés : donnée publique, interne, confidentielle ou strictement confidentielle. Une fois identifiées, ces données doivent être protégées.

Les obligations lors du transport des données

Sécurité : sécuriser les processus de transport des données

Le partage de données doit être effectué via des outils sécurisés, par des personnes habilitées et de façon confidentielle.

- **Etape 6 : Le stockage**

Les données constituent le patrimoine informationnel de notre entreprise. Posséder des données représente un réel avantage concurrentiel si son exploitation est pertinente. Les technologies de stockage vont alors jouer un rôle important dans le parcours de vie de la donnée.

Une donnée peut être stockée sur différents supports : téléphone, ordinateur, clé USB, disque dur, en pièce jointe d'un mail... Il existe de nombreux supports de stockage aux caractéristiques différentes.

Le choix du support de stockage doit se faire en fonction de différents paramètres, comme la capacité, la fréquence d'utilisation, les coûts, ou encore la durée minimale de disponibilité de ces données.

Les bonnes pratiques de stockage

Le stockage des données doit être sécurisé selon le niveau de classification de la donnée concernée.

Quatre niveaux de classification sont identifiés : donnée publique, interne, confidentielle ou strictement confidentielle. Une fois identifiées, ces données doivent être protégées.

Les obligations lors du stockage des données

Durées : respecter les durées de conservation des données

Le responsable des données est tenu de fixer une durée raisonnable de conservation des informations personnelles. En effet, une fois que l'objectif poursuivi par la collecte des données est atteint, il n'y a plus lieu de conserver les données. Elles doivent donc être supprimées. Cette durée varie selon les différents objectifs et doit tenir compte des éventuelles obligations légales à conserver certaines données.

Sécurité : sécuriser les processus de stockage des données

Les espaces de stockage doivent être sécurisés avec une habilitation pour modifier les données anonymisées et/ou cryptées.

- **Etape 7 : L'archivage ou la destruction**

L'archivage et/ou la destruction est la dernière, mais non la moindre, étape du parcours de vie de la donnée. Il est indispensable pour la donnée de passer par cette étape pour répondre aux enjeux business et/ou réglementaires qu'elle implique.

Certaines données personnelles peuvent, et dans certains cas, doivent, faire l'objet d'un archivage lorsqu'elles présentent encore un intérêt pour l'entreprise.

On distingue trois types d'archives :

- ✓ la base active (autrement appelée « archives courantes »),
- ✓ les archives intermédiaires (accès restreint, étape intermédiaire avant suppression),
- ✓ les archives définitives (données présentant un intérêt historique, scientifique ou statistique justifiant qu'elles ne fassent l'objet d'aucune destruction).

L'archivage et la destruction des données engendrent des coûts. Il est donc nécessaire de se fixer des objectifs relatifs au traitement des données pour se positionner sur de l'archivage ou de la destruction. Certaines données doivent être archivées car elles ont de la valeur pour une exploitation ultérieure.

Exemple des statistiques : grâce aux statistiques, il est possible de comparer des données entre elles sur différentes périodes pour en tirer des constats et mettre en place des actions. D'autres doivent être détruites dans le but de permettre un gain d'espace, de performance et de rapidité de réponse sur une base de données (il ne faut pas oublier qu'une donnée à un poids et peut donc être lourde).

Les bonnes pratiques d'archivage

L'archivage et/ou la destruction des données doit être sécurisés selon le niveau de classification de la donnée concernée.

Quatre niveaux de classification sont identifiés : donnée publique, interne, confidentielle ou strictement confidentielle. Une fois identifiées, ces données doivent être protégées.

Les obligations lors de l'archivage et la destruction des données

Sélection : veiller à archiver uniquement les données utiles

Lorsqu'un texte prévoit une obligation d'archivage, le responsable du fichier doit veiller à n'archiver que les données utiles au respect de l'obligation prévue, ou pour faire valoir un droit en justice.

Il doit, dès lors, opérer un tri parmi la totalité des données collectées pour ne garder que les seules données indispensables.

Durées : respecter les durées de conservation des données

Les données nécessaires pour répondre à une obligation légale ou réglementaire peuvent être archivées le temps nécessaire à l'accomplissement de l'obligation en cause. Les données ne doivent pas être conservées au-delà des durées de conservation légales et encore moins s'il n'est plus légitime de les conserver.

Sécurité : sécuriser les processus d'archivage et/ou de destruction

Des mesures techniques et organisationnelles doivent être prévues pour protéger les données archivées (destruction, perte, altération, diffusion ou accès non autorisés...). Ces mesures doivent assurer un niveau de sécurité approprié aux risques et à la nature des données.

Si des mesures de sécurité informatique sont indispensables, des mesures de sécurité physique doivent être prises, notamment lorsque des dossiers sont archivés sous format papier.

Lorsque l'archivage est confié à un sous-traitant, le responsable du fichier doit s'assurer que son prestataire présente des garanties suffisantes en matière de sécurité et de confidentialité des données qui lui sont confiées.

Quel que soit le type d'archive, la consultation des données archivées doit être tracée.

c. La qualité de la donnée

Comme exposé lors de la deuxième étape du cycle de vie de la donnée, les entreprises doivent rester vigilantes sur la qualité des données qu'elles utilisent tout au long du cycle de vie et prouver que chacun des critères de qualité est respecté à tout moment, de la collecte des données jusqu'à l'exploitation, en passant par le stockage. En effet, l'appréciation de la qualité des données peut se faire sur différents critères : l'exactitude, la fiabilité, l'exhaustivité, la pertinence, etc. (STEINBERG, 2016)

La donnée est facilement altérable au moment de sa création, lors des nombreuses manipulations, que ce soient des transformations automatiques ou manuelle, mais également au cours de son utilisation finale : c'est un actif fragile qu'il convient de protéger. D'après SRIVASTAVA (2014), nous pouvons rencontrer des erreurs sur les données, des doublons, des incohérences, des valeurs manquantes, incomplètes, incertaines, obsolètes, aberrantes ou peu fiables.

Pour cela, les entreprises doivent mettre en place des processus de sauvegarde, de conservation et de sécurisation afin de fournir une protection suffisante pour leurs données essentielles à l'exercice de leurs activités.

d. La marchandisation des données

La donnée est valorisable dès qu'il est possible d'en extraire une information utile. De ce fait, les compagnies d'assurance ont bien compris qu'il était dans leur intérêt de collecter le maximum de données sur leurs prospects ou clients afin de mieux segmenter, cibler et comprendre les comportements. Ainsi, il n'est pas rare qu'elles procèdent à des actions de monétisation des données pour mieux parvenir à les collecter. En échange de ses données personnelles, le client peut obtenir des réductions, des tarifs préférentiels, ou encore des cadeaux de bienvenue.

La numérisation du milieu économique dans lequel désormais les sociétés, couplée à l'expansion de réseau internet et de ses dérivés, ont décuplé l'intérêt des entreprises dans la captation de données. Les entreprises prennent conscience que les données peuvent leur

procurer un avantage concurrentiel sur leurs activités propres, mais également avoir une valeur marchande pour d'autres secteurs ou activités. Le marché de la donnée est né de ce constat, il est encore embryonnaire pour beaucoup de secteurs, mais ceux qui possèdent beaucoup de données ou sont en capacité d'en acquérir beaucoup par l'exercice de leurs activités peuvent entrevoir des relais de croissance très importants.

e. Les sources de données se diversifient

Les compagnies d'assurance collectent de grands volumes de données qu'ils ne détiennent pas en propre. Cette dernière tendance va s'accélérer dans les années à venir, afin de permettre de disposer de toutes les informations nécessaires à l'analyse de ses marchés, des attendus de leurs clients et prospects.

Il existe donc plusieurs typologies de données exploitées par les compagnies d'assurances, synthétisés dans le schéma ci-après :

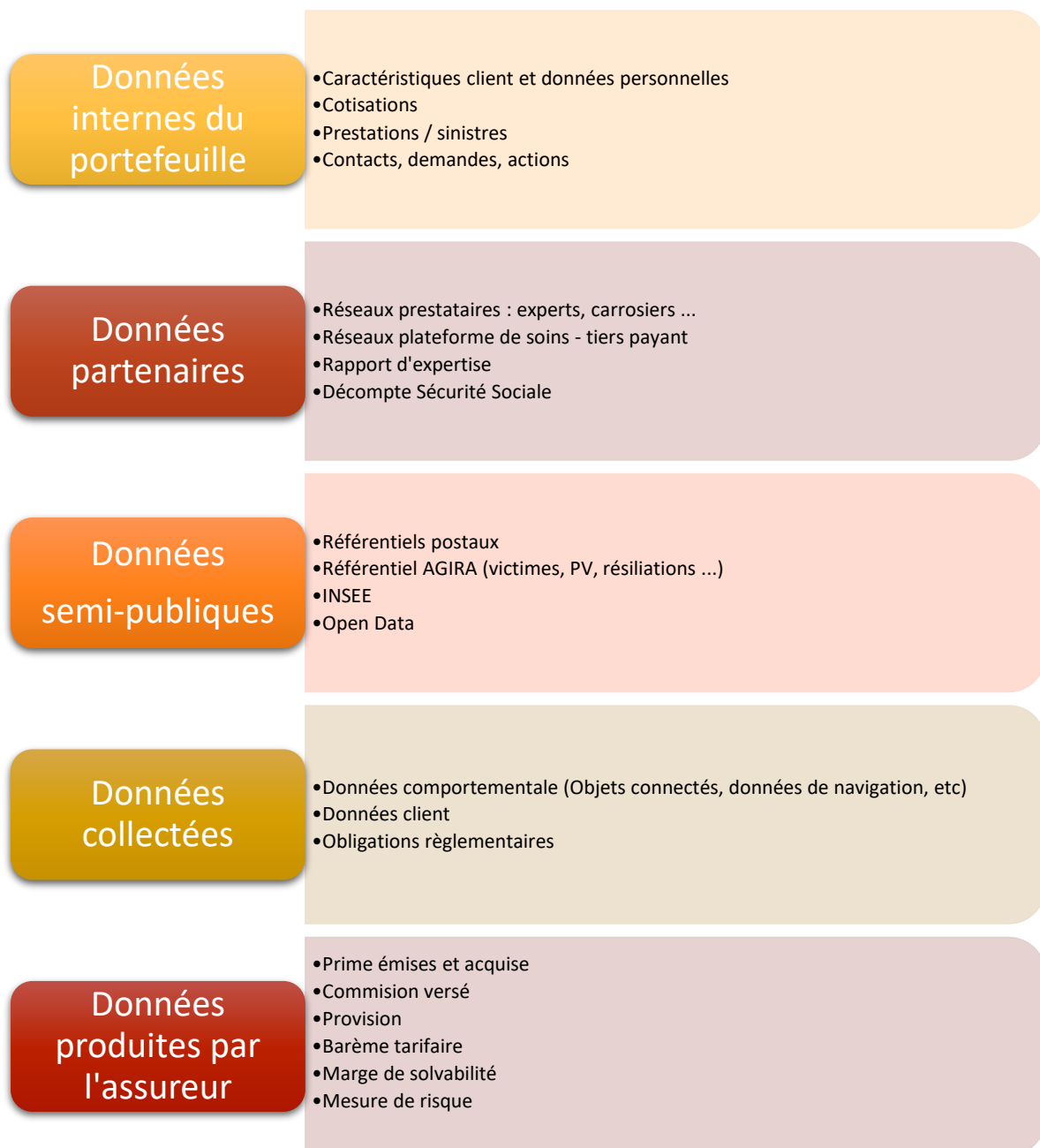


Figure 3 - Typologie des données utilisées par les assureurs

III. IDENTIFIER L'ÉCOSYSTÈME DU BIG DATA

a. Architectures et solutions Data

Définir les fonctions Big Data et de son écosystème n'est pas aisé tant la richesse et la diversité des modules et projets est importante. Les apports du Big Data et de ses outils d'information management et de business intelligence connexes peuvent être synthétisés de la manière suivante sur chacun des systèmes d'information :

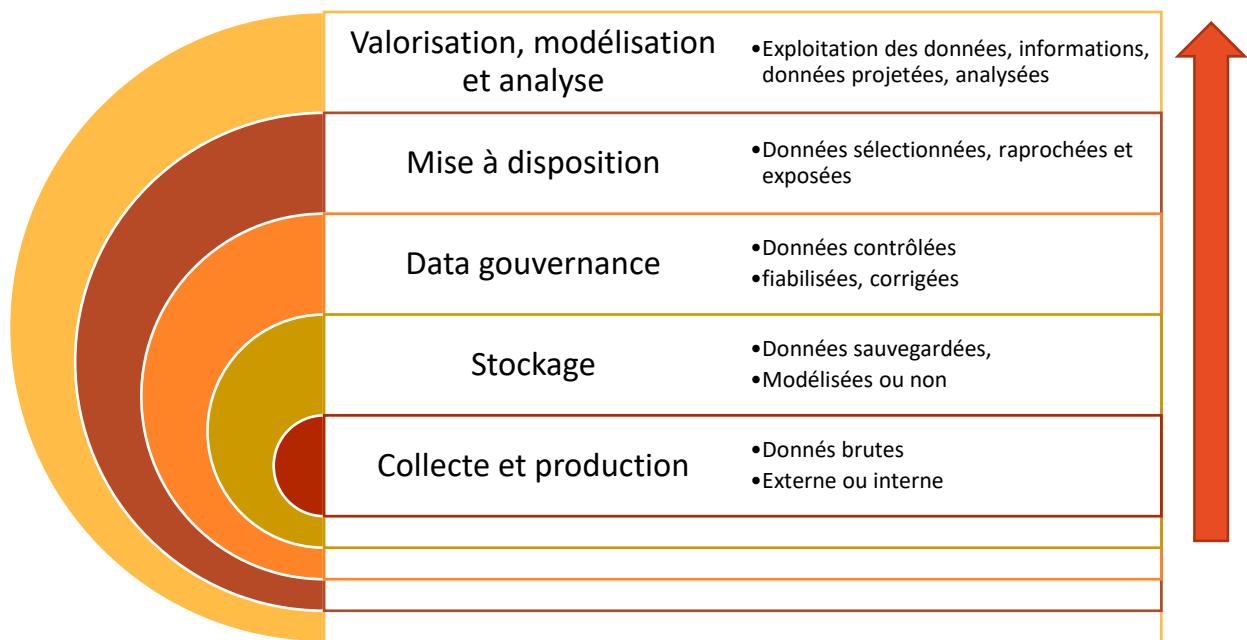


Figure 4 - Les 5 niveaux d'un système d'information

- **La collecte**

Le Big Data ne produit pas de données brutes, il les collecte et les affine. Il convient tout de même de se poser des questions fondamentales sur la nature et la qualité de ces données sources. En fonction de la nature des données recherchées et collectées, les systèmes doivent posséder des fonctionnalités diverses et spécifiques

La collecte de données externes et produites pose de nombreux problèmes directement liés à la nature de l'information collectée :

- **La pérennité des sources et des données collectées** : elle peut remettre en cause le fondement même de certains modèles et nécessite une veille spécifique et continue des sources de données ;
- **La fiabilité des données collectées** : une des spécificités des réseaux sociaux est l'absence de contrôle qualité et de recoupement des informations transmises. Une analyse sémantique ne suffit pas à identifier la véracité d'une information (ex : fake news) ;
- **La nature juridique des données collectées** : la confidentialité et le respect de la vie privée ;

- **Le stockage des données**

La fonction de stockage concerne les modalités de conservation des données à grandes échelles et à faible coût, tout en répondant aux exigences minimales en matière de disponibilité, d'accessibilité et de sécurité.

Outre les aspects de coûts informatiques liés au stockage, les modèles doivent intégrer les principes d'archivages et de gestion de temps d'exécution quel que soit la source ; de droit d'accès et de confidentialité ; et de sécurité.

- **La gouvernance des données**

Cette fonction comprend toutes les opérations de gestion des données et leur mise à disposition : elle fournit notamment les stratégies de mise à disposition de services (cloud, virtualisation des machines) la gestion des données référentielle et les métadonnées des modélisations. A cela s'ajoutent les tâches de contrôles et de correction de la qualité (QUANTMETRY, 2020).



Figure 5 - Les éléments clé de la gouvernance des données

Dans ce contexte, prendre en compte les impacts du Big Data est loin d'être neutre : nous serons souvent confrontés au problème du manque de données. Il s'agira de pallier ce déficit par la constitution de règles de corrections viables qui impactent le moins possible les modèles et les résultats.

- **La mise à disposition**

Les modèles et domaines de données gérés dans l'entreprise sont délimités par les fonctions métiers et les activités de l'entreprise. Le Big Data modifie substantiellement cette approche : l'entité peut désormais maîtriser des données externes à son activité pour mieux valoriser ses données internes. Ainsi, plusieurs niveaux d'utilisation des données Big data sont identifiés en fonction des cas d'usages et de leur implémentation.

En termes de solution, cette valorisation nécessite d'alimenter les systèmes de gestion des données internes avec des résultats ou des sous-ensembles de données Big Data.

- **Valorisation, modélisation et analytique**

Ce dernier niveau de la gouvernance des données sera présenté lors de notre cas d'usage chez l'assureur GENERALI.

- b. **Les algorithmes**

Les systèmes Big data de stockage et de calculs distribués, associés aux innovations sur le stockage, le calcul et la restitution des données, permettent d'étendre considérablement les capacités des systèmes statistiques. A ce titre, la révolution du Big Data est réhabilitée d'ancienne approche de statistique, d'analyse et de modélisation des données. Elle engendre donc de nombreux travaux de recherche permettant leur amélioration.

La diversité des algorithmes développés est telle aujourd'hui, qu'une des problématiques majeures devient le choix du bon modèle. Il existe deux grandes familles d'algorithmes permettant l'analyse et la modélisation des données :

- Les méthodes supervisées**

Les méthodes d'apprentissage supervisé sont des techniques d'apprentissage automatique dans laquelle on cherche à produire automatiquement un modèle à partir d'une base de données d'apprentissage. Dans cette approche, l'utilisateur sait ce qu'il cherche en définissant son modèle prédictif, il utilisera donc un algorithme de type « régression » s'il cherche à déterminer une variable quantitative et un algorithme de type « classification » s'il cherche à déterminer une variable qualitative.

- Les méthodes non supervisées**

Les méthodes d'apprentissage non supervisé, parfois nommé « clustering », sont utilisées lorsqu'aucun a priori n'existe sur les résultats et les sorties. Elles supposent de diviser un groupe hétérogène de données en sous-groupes de manière à ce que les données, considérées comme les plus similaires, soient associées au sein d'un groupe homogène. Au contraire, les données considérées comme différentes se retrouvent dans d'autres groupes distincts. L'objectif étant finalement d'extraire des connaissances organisées à partir de ces données.

Dans l'apprentissage non supervisé, un ensemble de données collectées est positionné en entrée. Ensuite un programme traite ces données comme des variables aléatoires et construit un modèle de densités jointes pour cet ensemble de données.

Les algorithmes font donc partie intégrante de l'univers du Big Data qui sans eux ne pourrait être exploité pleinement.

IV. L'IMPACT DU BIG DATA SUR LE SECTEUR DE L'ASSURANCE

a. L'environnement spécifique de l'assurance

Après avoir présenté les atouts et les limites du Big data, nous allons maintenant observer quels sont les incidences de celui-ci dans le secteur de l'assurance.

Dans son édito, J. François OUTREVILLE (2001) affirme qu'à la différence des autres industries, le cycle de production des assureurs est inversé. Habituellement, une société va investir dans de la recherche, acheter des matériaux, fabriquer un produit, supporter les coûts du stockage et de la distribution, avant de vendre sa production et d'être rémunérée, comme le montre le schéma ci-dessous.

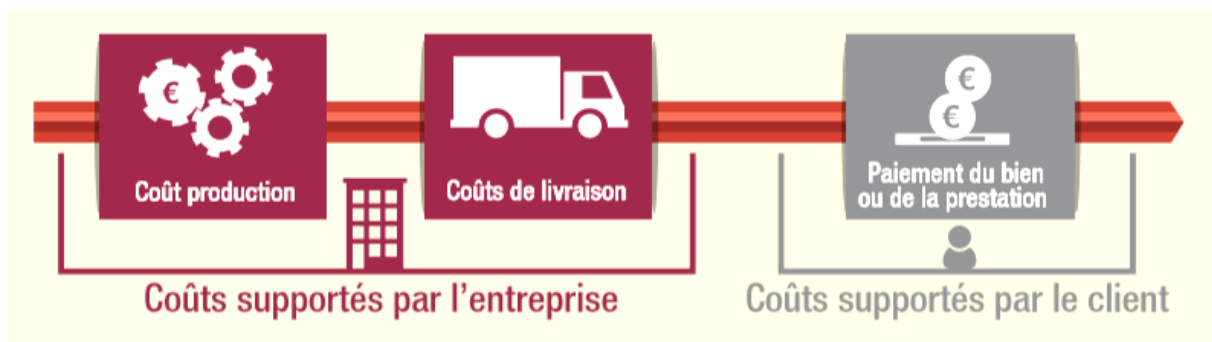


Figure 6 - Activité commerciale standard

Les compagnies d'assurances encaissent d'abord les primes versées par leurs clients, puis délivrent les prestations prévues au contrat. C'est ce que l'on appelle, avoir un cycle de production inversé.

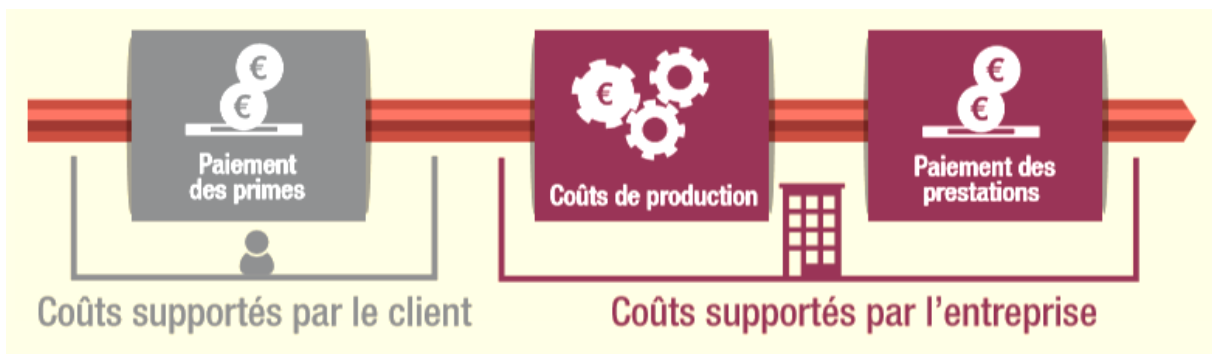


Figure 7 - Activité d'assurance

Il revient donc à l'assureur d'évaluer, avant que les sinistres ne surviennent, le coût des prestations à régler. Bien que le coût des prestations futures puisse être estimé à partir de l'historique du client, celui-ci demeure changeant. Cette incertitude est directement liée aux activités d'assurance, dont la finalité est de couvrir les risques, qui comme nous le savons sont par nature aléatoire et donc inconnue à l'avance.

Au-delà des risques à couvrir, l'ensemble des activités d'assurance dépend aussi plus ou moins du comportement des assurés. Par exemple, en épargne, la tendance d'un client à

racheter un contrat d'assurance-vie peut dépendre aussi de facteurs externes qui ne sont pas directement liés à l'évolution de l'épargne dans le contrat, notamment :

- L'environnement concurrentiel ;
- La situation personnelle de l'assuré ;
- L'évolution des autres types de placements dans lesquels l'assuré peut arbitrer son épargne.

Par conséquent, il est très important pour les compagnies d'assurance de connaître et d'appréhender, tant les risques qu'ils couvrent que les comportements des clients.

b. Les solutions de Big Data : Quels impacts en assurance ?

Nous allons maintenant présenter les principaux pôles dans l'assurance, qui sont plus ou moins impactés par les solutions Big Data :



Figure 8 – Principaux pôles d'assurance impactés par le Big Data

▪ **Produit & Garanties**

Le Big Data a un impact important dans le secteur de l'assurance, car la donnée en est l'essence. Les sources de données qui peuvent être utilisées pour concevoir de nouveaux prix et contrôler les risques liés aux produits ou aux garanties sont de plus en plus nombreuses. De même pour le nombre de variables explicatives de la sinistralité permet de réajuster la tarification, le but est de proposer aux clients des prix plus compétitifs et des garanties personnalisées, plus adaptées à leurs profils.

Les cas d'utilisation les plus pertinents aujourd'hui, concernent les produits d'assurance santé et automobile. Pour exemple, dans l'assurance santé, des initiatives sont prises comme chez GENERALI, pour valoriser la pratique sportive, facteur de prévention de risque maladie aidé par des objets connectés collectant des données de « bien-être ». En concevant des garanties qui prennent en compte les données sociales, météorologiques, ... le Big Data aura un impact plus important.

▪ **Gestion des risques**

Comme présenté précédemment, les compagnies d'assurances doivent prendre en compte différents facteurs risques. Certains de ces risques sont plus ou moins maîtrisés. Philippe SIMON, le Président de la Cegedim Insurance Solutions, explique : « *Tous les acteurs de l'assurance sont peu ou prou confrontés à la nécessité de réduire leurs coûts, afin d'améliorer leur S/P¹. [...] réduire le coût du risque représente un levier d'action trois fois plus puissant que la diminution des frais de gestion* » (LIMOGE, 2017). Ainsi, pour les risques moins maîtrisés comme la résiliation ou le rachat d'un contrat, les solutions Big data offrent donc la possibilité de modéliser ceux-ci, permettant un meilleur contrôle dans la gestion des risques.

▪ **Lutte Anti-Fraude**

Toujours dans la réduction des coûts des risques, le Big Data intervient aussi dans la lutte anti-fraude. D'après Philippe SIMON, il permet ainsi de « ne pas payer à tort » en détectant les comportements amoraux, aspirant à maximiser les garanties. (LIMOGE, 2017).

Aujourd'hui, de nombreuses méthodes de détections de profils à risque sont proposées par le Big Data. Selon une étude publiée en 2007 par le cabinet de consultants PAC, 63% des compagnies d'assurances affirment utiliser les solutions Big Data dans la lutte anti-fraude. Cependant, il est vrai que les acteurs de la santé, plus particulièrement les compagnies d'assurance se retrouvent confronté à des problèmes de formations, de transformation organisationnelle ou encore de stratégie.

▪ **Pilotage Technique et Financier**

Aujourd'hui, le pilotage technique et financier est de plus en plus lié aux solutions Data. De nombreuses entreprises s'en remettent et disposent maintenant d'une visibilité-client idéal leur permettant de d'établir de nouvelles stratégies financières.

Avec les différentes solutions data, les compagnies d'assurances sont maintenant capables d'assurer une veille stratégique et financière, et de créer des indicateurs de performance nécessaire au pilotage.

▪ **Marketing**

Présent dans tous secteurs d'activités confondus, les cas d'utilisation les plus courants du Big Data à ce stade sont liés à l'amélioration de la connaissance client. L'objectif est de mieux prévoir leur comportement d'achat afin de leur proposer au meilleur moment des produits ou des services au plus proche de leurs besoins.

Les solutions Big Data ont aussi une utilisation discrète et ponctuelle, elles sont nécessaires au provisionnement et à la tarification.

¹ S/P : Rapport entre le montant des sinistres et des cotisations permettant de mesurer la rentabilité d'un contrat d'assurance

➤ Provisionnement

Les provisions techniques et mathématiques correspondent à l'ensemble des réserves constituées par les compagnies d'assurance pour honorer leurs engagements : pensions de décès, rentes et versements d'assurance vie (PREVISSIMA, 2020).

La détermination des provisions peut être perçue comme un processus continu. Cependant, la mise en œuvre des approches Big Data n'impacte pas les méthodes de provisionnements mais les paramètres utilisés dans la détermination des provisions qui sont, eux revus ponctuellement. (KHORDJ et al., 2017).

➤ Tarification

Les revues tarifaires et de garanties sont des processus ponctuels bien récurrents. La connaissance client, permise par les approches Big Data et la volumétrie des données qu'elles exploitent, sont de nature à affiner ces processus. Les objectifs de la tarification sont :

- De proposer des garanties plus adaptées à la typologie des clients
- De mettre en œuvre des tarifs adaptés aux risques que présentent les clients.

Ainsi, les solutions Big Data interviennent à l'aide des données relatives à l'environnement dans lequel évoluent les assurés potentiels, ou encore les données relatives aux assurés comme la fréquence des sinistres ou leur coût moyen.

c. Des impacts métiers pour l'assureur

L'avènement du Big Data va entraîner le traitement de données complémentaires ainsi qu'une forte valorisation des données directement détenues par les assureurs. Cette valorisation, au travers d'algorithmes d'exploitation de données, induira des traitements complémentaires à ceux actuellement réalisés. Ces travaux nécessitent l'émergence de profils de data scientists maîtrisant :

- Les solutions de Business Intelligence ;
- Des compétences pointues en algorithmique ;
- Des connaissances métier pour contribuer à l'identification des usages autorisant un gain de valeur pour l'entreprise, condition sine qua non au lancement de projet autour du Big Data.

Actuellement, les profils disposant du triptyque de compétences attendues des data scientist sont rares mais des formations sont progressivement mises en place par les écoles et les universités. MCKINSEY estime, tous secteurs confondus que la demande en profils de type data scientist pourrait atteindre 1.5 million de personnes dans le monde dans cinq ans (PIATETSKY, 2018).

La mission des data scientists consiste à programmer des algorithmes statistiques sur les nouveaux outils et technologies du Big Data pour identifier et interpréter des indicateurs, issues de données volumineuses. La seconde étape consiste à les présenter sous une forme compréhensible pour les interlocuteurs métiers avec toute l'interactivité nécessaire aux analyses et aux prises de décision. Les data scientist doivent faire preuve d'ouverture, de

créativité et être également très à l'écoute des innovations technologiques et statistiques susceptibles d'être mises à disposition. Ils doivent par ailleurs organiser et réaliser une veille autour de données externes pour identifier de nouvelles sources de valorisation de patrimoine données de l'entreprise.

Si l'émergence de nouveaux usages et de nouveaux profils peut être de nature à inquiéter certaines professions, notamment les actuaires, il y a fort à parier qu'au moins à court terme, l'impact sur l'organisation et les prérogatives des fonctions existantes sera limité. Il conviendra surtout pour ces fonctions de s'adapter aux nouveaux usages et d'intégrer pleinement les apports et expertises de ces nouveaux métiers.

Le tableau ci-dessous récapitule les prérogatives nécessaires en assurance pour deux professions : les actuaires et les data scientist. Il montre bien les spécificités, mais aussi les convergences pouvant faire de ces deux fonctions un duo gagnant pour relever les défis à venir.

Actuaire	Data Scientist
Principales prérogatives	
<ul style="list-style-type: none"> ✓ Tarification ✓ Provisionnement ✓ Réalisation d'inventaires ✓ Respect des contraintes règlementaires et prudentielles ✓ Mise en œuvre et utilisation de modèles de projection 	<ul style="list-style-type: none"> ✓ Veille sur les sources de données ✓ Veille sur les solutions Business Intelligence ✓ Mise en œuvre et utilisation d'algorithmes prédictifs ✓ Identification des comportements clients ✓ Optimisation de la valeur clients
Contribution à la gestion de risques Gestion de la qualité des données	

Figure 9 - Complémentarité des actuaires et des data scientist

V. PASSER DE LA THEORIE A LA PRATIQUE

Comme nous l'avons vu précédemment, nous pouvons attester de l'ubiquité du Big Data dans le secteur assurantiel. Si bien que le PDG de McKinsey Digital Group affirme que dans ce secteur, la création de valeur-client y est beaucoup plus importante que dans toute autre industrie. Néanmoins, est-il optimal pour la mise en œuvre de projet data ? Dispose-t-il de tous les critères nécessaires pour sa réalisation ? Pour répondre à ces interrogations, il serait convenable de comprendre comment mener un projet data et qu'elles sont les différentes méthodologies permettant d'y arriver.

a. Comment concevoir un projet data ?

Avant de dresser un portrait des différentes ressources utilisées lors d'un projet Big Data, il paraît fondamental de poser quelques jalons nécessaires pour sa compréhension. Nous allons donc définir qu'est-ce qu'un projet et quels en sont les acteurs.

i. Qu'est-ce qu'un projet ?

Etymologiquement, le mot « projet » provient du latin *projectum* et possède plusieurs significations. Il peut se définir comme un ensemble d'activités coordonnées et liées visant à atteindre un objectif spécifique, qui est généralement atteint sur une période de temps prédéterminée tout en respectant un budget. Dans le langage courant, il peut également être utilisé comme synonyme de plans, de procédures et d'idées.

Une autre définition, plus en adéquation avec la gestion de projet, nous est donnée par la norme NF EN ISO 9000 : le projet est un « *processus unique, qui consiste en un ensemble d'activités coordonnées et maîtrisées comportant des dates de début et de fin, entrepris dans le but d'atteindre un objectif conforme à des exigences spécifiques, incluant des contraintes de délais, de coûts et de ressources* » (Norme NF EN ISO 9000, 2005).

De cette définition, nous pouvons retenir 4 points essentiels :

Processus unique : La méthode projet appartient au domaine de l'innovation et de la créativité. Nous devons alors faire la distinction entre les activités « projets » non récurrentes et les activités « opérationnelles » régulières.

Ensemble d'activités coordonnées et maîtrisées : Le projet comporte de nombreuses tâches de nature différente, qui doivent être maîtrisées tout au long du processus.

Dates de début et de fin : Le projet a une durée, il a un commencement et une fin. Il est très important de maîtriser

Exigences spécifiques : Les exigences spécifiques sont exprimées dans un cahier des charges.

Ressources : le projet mobilise des ressources humaines et matérielles.

Dans son article « De la pédagogie de projet à la conception de projet » (RAYNAL et al., 2008), S. Raynal complète cette définition en précisant qu'un projet nécessite obligatoirement un cadre méthodologique spécifique. En effet, réaliser un projet en équipe est généralement plus facile à dire qu'à faire. Il existe donc des méthodes de gestion de projet nous permettant de

réaliser celui-ci avec discipline, augmentant ainsi l'efficacité de l'équipe. L'organisation du projet est alors optimale. Ainsi, la méthodologie de projet permet donc de guider, de planifier, d'initier et mettre en œuvre celui-ci.

Cependant, d'après le « Top 13 des méthodes de gestion de projets » (AZENDOO, 2016), il existe différentes méthodes pour différents types de projets. Chaque modèle de gestion de projet n'est pas forcément adéquat pour les différentes tâches que comporte celui-ci. Nous allons maintenant dresser le portrait de ces différentes méthodologies pour déterminer lequel serait optimal pour le secteur assurantiel.

Il existe une multitude de méthodologies de gestion de projet, nous pouvons citer : la méthode du chemin critique, la méthode PERT, la méthode CRISTAL, Six Sigma et bien d'autres, mais nous ne vous présenterons que deux méthodes, la méthode dite traditionnelle et la méthode Agile, car ces deux méthodes sont aujourd'hui, celles les plus utilisées quel que soit le secteur d'activité.

b. Les méthodes Agile

Les méthodes « Agiles » sont particulièrement adaptées à la mise en œuvre des projets Big Data. Elles consistent en un ensemble de pratiques conçues pour pallier aux difficultés rencontrées dans les cycles de développement plus classiques et encore omniprésents dans les structures d'assurances (Cascade, cycle en V etc.). L'objectif ici n'est pas de couvrir l'ensemble de la méthodologie « Agile » mais d'en expliquer les grandes lignes afin de montrer comment celles-ci se prêtent volontiers à l'implémentation d'un prototype Big Data.

c. Les méthodes traditionnelles

Les méthodes traditionnelles prônent un enchaînement séquentiel des différentes phases d'un projet, depuis les spécifications jusqu'à la validation du système, selon un planning préétabli. Elles visent à mieux prédire la façon dont les choses « devraient » se passer. Malheureusement, cette vision rassurante est bien loin de la réalité des projets. Les activités d'ingénierie ne sauraient se succéder strictement sans qu'aucun changement en viennent perturber un planning qui n'a souvent de durée de vie que le temps de la concevoir.

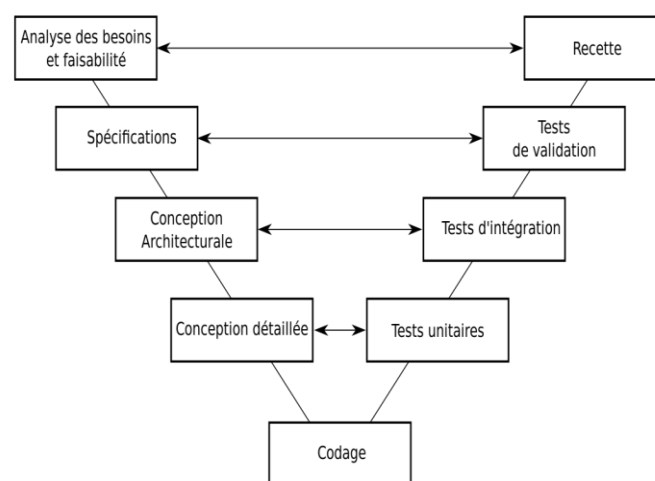


Figure 10 - Méthodologie Cycle en V

Une majeure partie des projets réalisés en suivant les méthodes traditionnelles, connaît des retards de livraison, des dépassements budgétaires voire des échecs partiels ou totaux pour n'avoir pas satisfait les besoins des utilisateurs.

Les méthodes agiles préconisent l'adoption d'un cycle court, itératif et incrémental permettant à une équipe de s'adapter au contexte ainsi qu'aux changements qui ne manquent pas de survenir au cours d'un projet. L'implication du client dans le développement permet à celui-ci de suivre au plus près la construction de l'application ou du projet en cours de développement, écartant ainsi tout « effet tunnel ».

La définition d'objectifs à court terme permet de maintenir une pression constante mais supportable sur l'équipe, alors qu'au début d'un cycle en V, chaque acteur peut avoir l'impression de disposer suffisamment de temps devant lui et subit finalement une très forte pression à l'approche de la livraison.

Le produit livré est opérationnel et de bonne qualité parce qu'il est souvent testé. Il est doté de la seule documentation strictement nécessaire, et répond à coup sûr aux vrais besoins des utilisateurs puisqu'il est régulièrement soumis à leur feedback.

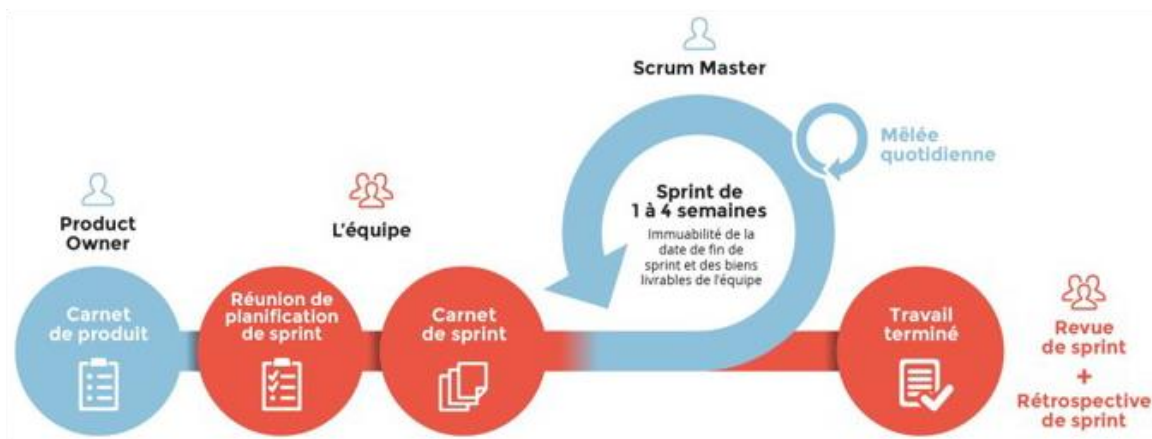


Figure 11- Méthode Agile [16]

Forts de ces principes, il peut être constaté qu'une organisation, un département, une équipe ou un projet peuvent adopter l'agilité avec succès.

L'agilité peut également améliorer les résultats déjà obtenus et faciliter la résolution de bon nombre de difficultés vécues. Elle va amener les personnes impliquées à mieux collaborer, prendre du recul sur l'application en priorisant les actions. (GREGORY, 2017)

d. L'utilisation des méthodes Agile pour les projets Big data

Les méthodes agiles se prêtent naturellement aux projets de type Big Data et, en général, à l'ensemble des projets de type recherche et développement, lorsqu'un objectif est défini mais que la manière d'y parvenir est encore incertaine. La définition de cas d'utilisation et leur mise en application nécessite constamment des aller-retours entre les différents acteurs. Le caractère incertain des données manipulées et des corrélations qu'elles sont susceptibles de révéler, nécessite un modèle de développement itératif, incrémental et adaptable aux besoins métiers.

VI. CERNER LES ENJEUX REGLEMENTAIRES ET JURIDIQUES

a. Le Règlement Général sur la Protection des Données

D'après le site internet de la CNIL, le 25 mai 2018, le RGPD est entré en vigueur. Cette nouvelle réglementation européenne signifie que les entreprises auront l'obligation d'obtenir le consentement explicite des individus pour stocker et utiliser leurs données personnelles. (CNIL, 2018)

Finalement, le RGPD devient une opportunité pour les compagnies d'assurance, leur permettant d'améliorer leur relation avec leurs clients et de se distinguer autour des valeurs de respect pour établir un véritable sentiment de confiance (CNIL, 2018).

b. Une protection des données règlementée par la CNIL

Nous pouvons définir la CNIL comme étant « la Commission Nationale de l'Informatique et des Libertés », chargée de la gestion des données personnelles. Elle s'assure de la conformité de l'utilisation des données auprès des particuliers et des professionnels. » (CNIL, 2020).

Sa mission principale s'articule autour de quatre piliers :

1- INFORMER ET PROTEGER

La CNIL intervient régulièrement auprès des professionnels et les particuliers afin de prévenir ceux-ci à l'aide d'outils pédagogiques pratiques sur les nouvelles utilisations du numériques.

Elle sensibilise donc les usagers sur l'utilisation de la technologie de protection de la vie privée, y compris la technologie de cryptage des données.

2- ACCOMPAGNER / CONSEILLER

Le contrôle des données à caractère personnel implique une variété d'outils, tous destinés à assurer la conformité des organisations. Elle permet d'affirmer que le processus d'anonymisation des données personnelles répond à son objectif de publication et de réutilisation.

3- CONTROLER ET SANCTIONNER

Sur la base des problèmes d'actualités, des problèmes majeurs constatés et des plaintes déposées auprès de la CNIL, un plan de contrôle est élaboré. Lors d'une inspection, la CNIL a le pouvoir de :

- Pénétrer dans tous les locaux commerciaux
- Accéder à la documentation, programmes informatiques et autres
- Recueillir toutes les informations utiles et écouter les opinions de chacun

Après l'inspection, le président de la CNIL peut décider de le notifier formellement des mises en conformité. Elle peut avoir recours à des sanctions :

- Avertissements publics ;
- Sanctions financières ;
- Publication des sanctions ;
- Retrait de l'autorisation d'exploitation des données ;

4- ANTICIPER

Dans le cadre de ses activités innovantes et tournées vers l'avenir, la CNIL a pour but premier de détecter et analyser les nouveaux usages de la technologie susceptibles d'avoir un impact significatif sur la vie privée.

Sa mission est de réfléchir aux enjeux éthiques et sociaux engendrés par le développement du numérique.

PARTIE II - GENERALI ASSURANCE

I. HISTORIQUE

Société anonyme d'assurance, GENERALI, anciennement nommée Assicurazioni GENERALI Austro-Italiche a été fondé en 1831 à Trieste (Italie) par Giuseppe Lazzaro Morpurgo et quelques entrepreneurs dans le but de concurrencer des compagnies d'assurances émergentes en Lombardie et Vénétie. En quelques années, GENERALI conquiert l'Empire autrichien et les grands ports européens en s'installant dans les grandes villes comme Marseille en 1834 et Bordeaux en 1832.

Dans un contexte historique particulier, elle a su imposer sa vision au niveau internationale. Face à la Seconde Guerre mondiale, GENERALI a perdu des actifs et des succursales dans les pays touchés par l'Union soviétique. Le groupe se redresse rapidement dans les pays d'Europe de l'Ouest et se concentre sur la conquête de nouveaux marchés en Amérique latine. Cette période de fort développement économique a conduit à l'émergence de nouveaux risques, et donc au développement de l'assurance.



Figure 12 - Implantation de GENERALI dans le monde aujourd'hui

Aujourd'hui, le groupe GENERALI est présent dans plus de 60 pays et compte plus de 500 entreprises à l'international. En comptant plus de 80 000 employés, elle possède un réseau de distribution de plus de 100 000 employés dans le monde, avec près de 65 millions de clients. En Europe, GENERALI est la première compagnie d'assurance vie et la troisième compagnie d'assurance générale. A l'international, selon l'agence de notation AM Best, GENERALI se classe deuxième en primes émises et cinquième en actifs.

II. GENERALI FRANCE

Aujourd'hui, GENERALI France est l'une des principales compagnies d'assurance en France. Avec un chiffre d'affaires de plus de 13,3 milliards d'euros en 2019, il fournit des solutions d'assurance et de gestion d'actifs à 7,4 millions de personnes et 750 000 professionnels et entreprises. GENERALI se différencie en se positionnant et en proposant des services innovants dans tous les domaines de l'assurance.

Ces principales branches d'activités sont la Vie (société Generali Vie) et l'activité Dommages et Responsabilités (société GENERALI IARD)

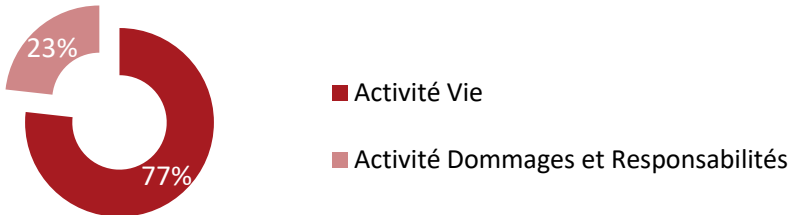


Figure 13 - Répartition de l'activité chez GENERALI

III. PRESENTATION DU SERVICE

Au sein de la Direction Technique Assurance, j'ai été affecté dans l'équipe Prévoyance & Santé individuelle et Professionnelle.

Rattaché au Bureau d'Etudes Techniques Vie, l'entité Prévoyance & Sante Ind. Prof. a pour vocation de concevoir et actualiser les produits d'assurance, de définir les règles de souscription et de gestion et de leurs tarifs. Elle a aussi pour mission de suivre les résultats ainsi que la rentabilité de chaque produit, tout en sécurisant la gestion des portefeuilles. Elle participe aussi à la mise en place d'outils et d'indicateurs de valorisation des clients.

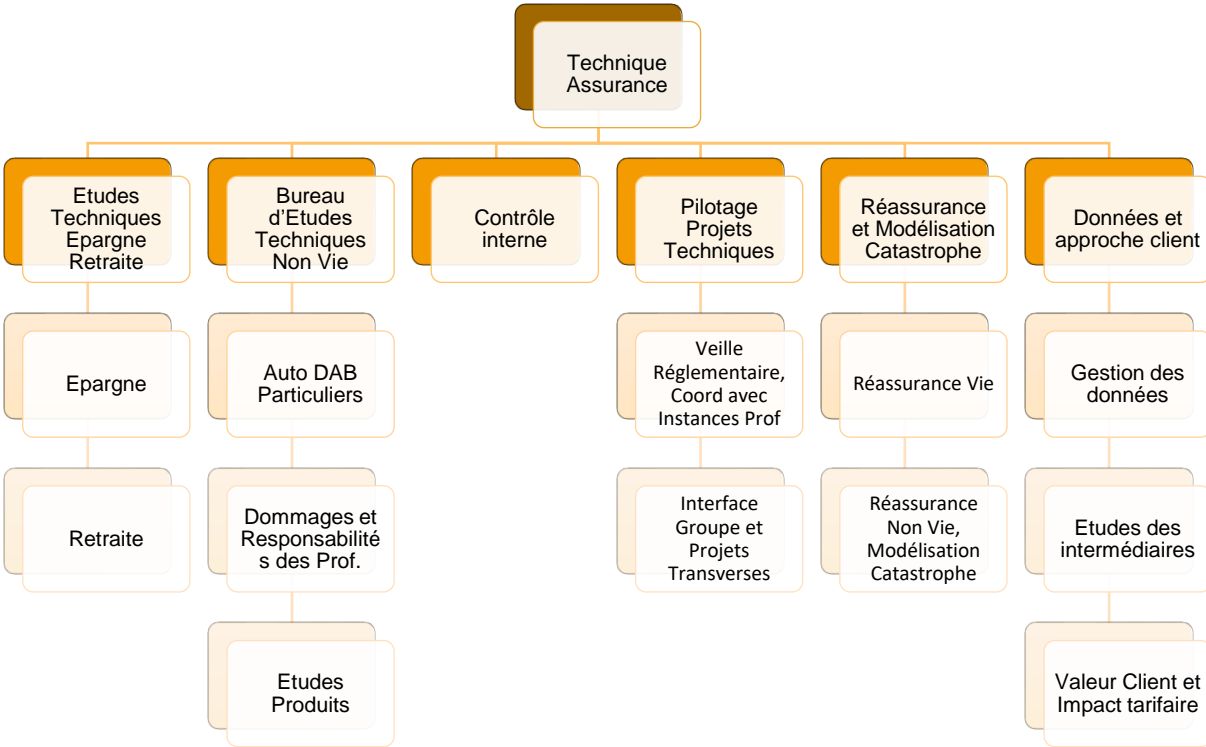


Figure 14 - Organigramme de la Technique Assurance chez GENERALI Assurance

L'ASSURANCE SANTE & PREVOYANCE

GENERALI garantit une gamme très large et propose à ses clients des produits d'assurance dommages plus généraux et des assurances de personnes. Ces produits sont principalement destinés à des particuliers, des professionnels, des collectivités ou des entreprises, professionnels du bâtiment ou encore aux professionnels de l'automobile.

L'infographie ci-dessous synthétise l'ensemble des produits proposés par GENERALI. Nous nous intéresserons donc aux produits de la Prévoyance & Santé.



Figure 15 - Les différents produits d'assurance chez GENERALI

I. DEFINITION DE L'ASSURANCE SANTE

L'assurance santé vise à protéger un assuré des conséquences économiques liées à la survenance de risques conduisant à un acte ou une intervention médicale. Une autre définition plus technique de l'assurance santé, nous est donnée par le « Guides des assurances et des mutuelles santé » :

« L'assurance santé est un système économique par lequel un ou plusieurs assureurs garantissent – après signature d'un contrat d'assurance (appelé « contrat d'assurance santé ») engageant le ou les assureurs avec un souscripteur, et selon les conditions figurant au contrat – un ou plusieurs assurés contre les conséquences financières entraînées suite à la réalisation d'un événement dommageable ayant porté atteinte à l'intégrité physique (ou parfois psychologique) dudit ou desdits assurés, et nécessitant une intervention d'ordre médical pour tenter de rétablir le ou les sinistrés dans la situation où ils se trouvaient avant la survenance du sinistre » (WILHELM, 2020).

Nous pouvons retenir de cette définition :

- ✓ L'assurance santé offre une protection financière pour prévenir les risques qui affectent la santé de l'assuré, plutôt que les risques susceptibles d'affecter ses biens ;
- ✓ Il faut faire la distinction entre les souscripteurs et les assurés. Le souscripteur est le signataire d'un contrat d'assurance santé, qui s'engage à régler les primes d'assurances à la compagnie d'assurance. L'assuré, lui, est considéré comme « consommateur » et bénéficiaire des services fournis par l'assureur. Ainsi, dans un contrat d'assurance santé, il y a le souscripteur qui est bénéficiaire, mais il peut y avoir aussi plusieurs autres bénéficiaires bénéficiant tous des mêmes garanties prévues par le contrat.

Par définition, l'assurance santé ne rembourse les frais médicaux que dans les circonstances susmentionnées et dans les limites prévues du contrat. Cela signifie qu'il est nécessaire de déterminer clairement les besoins en assurance santé des bénéficiaires du contrat pour que les compagnies d'assurance sélectionnent et proposent des solutions d'assurance les plus adaptées.

a. Comment fonctionnent les remboursements en santé ?

Lors de la survenance d'un sinistre, autrement dit lorsque la compagnie d'assurance doit rembourser les frais de santé d'un assuré, il existe plusieurs niveaux de remboursement :

- **1^{ère} tranche** : Base de Remboursement par la Sécurité Sociale (BRSS). Elle prévoit un remboursement sous forme de taux appliqué sur une base de tarif fixée.
- **2^{ème} tranche** : Remboursé par la compagnie d'assurance. Selon le niveau de garantie du contrat d'assurance santé, l'assuré bénéficiera d'un remboursement plus ou moins important. Le remboursement est exprimé en euros ou en pourcentage de la base de sécurité sociale :
 - Le remboursement en euros : Il s'agit d'un montant total déterminé chaque année pour payer les frais médicaux sur certaines garanties. Lorsque les frais de santé dépassent le forfait stipulé dans le contrat d'assurance santé, les frais supplémentaires sont à la charge de l'assuré.
 - Le remboursement en pourcentage de la base de la Sécurité sociale : L'assureur indemnise un certain pourcentage de la base de remboursement de la Sécurité sociale.
- **3^{ème} tranche** : Reste à charge par l'assuré. Comme annoncé dans la 2^{ème} tranche, en fonction de la part prise en charge par le contrat d'assurance santé, il peut rester une partie de la prestation à la charge de l'assuré. Cependant, au 1^{er} janvier 2020, le gouvernement a lancé une nouvelle réforme : le Reste A Charge zéro (RAC 0), plus connu sous le nom de « 100% Santé ». Comme son nom l'indique, cette réforme consiste à la suppression du reste à charge après

le remboursement de la Sécurité Sociale et des complémentaires santé : favorisant l'accès aux soins à de nombreux ménages.

Par exemple, pour une couronne dentaire à 2 000 €, la base de remboursement est de 107,50 €. Le taux appliqué est de 70%. Pour une garantie à hauteur de 100% BRSS, le remboursement de la Sécurité Sociale sera donc de 75,25 €. Si l'assurance propose un remboursement de de XXX% de BRSS, elle remboursera 547.25€ à l'assuré. Il lui reste 1 337.50€ à charge.

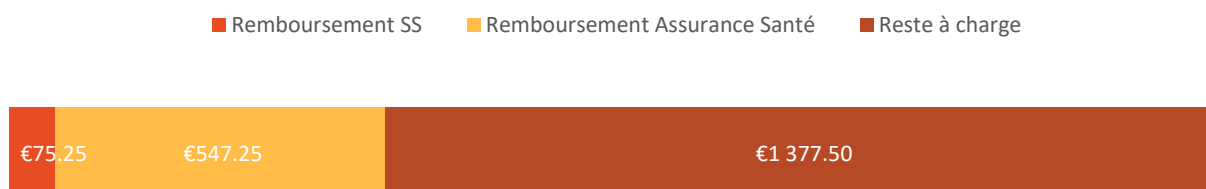


Figure 16 - Exemple de remboursement

b. Plusieurs niveaux de couverture

D'après Noëlle HERMAL, le remboursement des frais de santé est assuré par la Sécurité Sociale et par le contrat d'assurance santé souscrit par l'assuré. Cependant, les contrats d'assurance santé fournissent plusieurs niveaux de garanties plus ou moins protecteurs.

- **Le contrat basique : Ticket modérateur**

Le Ticket Modérateur prévoit un remboursement à 100% de la BRSS dans tous les postes de soins. En d'autres termes, pour les frais d'optiques ou dentaires, le remboursement sera calculé sur la base du remboursement de la sécurité sociale.

- **Le contrat intermédiaire**

Certaines compagnies d'assurances proposent à leurs assurés des contrats dits « intermédiaires ». Le montant du remboursement peut varier de 150% à 250% de la BRSS. Ces niveaux de garantie couvrent partiellement ou totalement les dépenses en frais médicaux pour des actes comme l'optique, le dentaire, l'hospitalisation ou encore les soins courants.

Ces niveaux de contrat couvrent davantage les frais médicaux, même ceux qui ne sont pas remboursés par la sécurité sociale. Le taux de remboursement est plus élevé que le contrat de base. Pour les postes de santé à faible remboursement de la sécurité sociale, l'entraide avec garanties intermédiaires prévoit généralement un forfait. (HERMAL, 2019)

Pour ce niveau de garanties, GENERALI propose des formules adaptées aux besoins des assurés : formule « familles », « seniors », etc.

- **Le contrat haut de gamme**

Ce niveau de contrat propose une compensation allant de 300 à 800%. Ils prévoient également le remboursement de toutes les actions non payées par la Sécurité sociale. Ces contrats sont très complets et s'adressent généralement aux personnes âgées ou aux personnes qui ont une forte demande de remboursement des frais médicaux. Ce sont aussi des contrats assez chers, qui s'adresse à ceux qui veulent être bien protégés même si le prix est élevé.

En somme, les postes principaux de remboursements des contrats d'assurance santé sont : l'hospitalisation, le dentaire, l'optique et les soins courants, qui regroupe l'ensemble des consultations chez les médecins généralistes et les spécialistes, (FRUCHARD, 2020).

II. DEFINITION DE L'ASSURANCE PREVOYANCE

De nombreux aléas peuvent affecter la vie d'un individu : la maladie, des accidents ou encore le chômage, ont un impact considérable et parfois irréparable sur le niveau de vie d'un assuré. Les contrats de prévoyance permettent aux individus d'anticiper ces risques afin de compenser les pertes financières ou de gérer efficacement le statut de dépendance.

Le droit français définit la prévoyance comme « *opérations ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage.* » (Loi EVIN,1989).

La législation encadre aussi les institutions proposant des contrats de prévoyance. Sont donc autorisés à proposer ce type de couverture :

- a. Entreprises régies par le code des Assurances ;
- b. Institutions de prévoyance relevant du titre III du livre IX du code de la Sécurité sociale ;
- c. Institutions de prévoyance ;
- d. Mutuelles relevant du code de la Mutualité ;
- e. Organismes visés aux articles L. 644-1 et L. 652-1 du code de la Sécurité Sociale.

En tant qu'entreprise régie par le code des Assurances, GENERALI est autorisé à proposer des contrats de prévoyance.

a. Les principaux types de contrats de prévoyance

Nous pouvons tous être atteint d'une maladie, faire une mauvaise chute, ou encore une blessure. Nulle ne sait si un jour nous y serons confrontés. Pour anticiper ces risques, les compagnies d'assurances proposent des solutions de prévoyances composées de garanties répondant à ses besoins spécifiques. Il existe donc plusieurs types de garanties garant à différents types de risques.

- **Garantie des accidents de la vie**

La garantie accidents de la vie (GAV) sert à protéger l'assuré, et éventuellement sa famille, des conséquences des accidents de la vie quotidienne. L'assureur indemnise l'assuré victime d'accident, si le responsable n'est pas identifié ou si l'assuré est lui-même le responsable. Les contrats proposés par les assureurs diffèrent, certains éléments de la garantie peuvent être déjà couverts par d'autres assurances (sécurité sociale, complémentaire santé, carte bancaire...) (Direction de l'information légale et administrative, 2019).

La garantie peut varier en fonction de l'assureur et du montant des cotisations.

- **Assurances décès et obsèques**

Le décès est l'un des risques pris en compte dans la définition de la prévoyance. Il existe 2 types de contrats répondants à ces besoins :

- L'assurance Décès permet à la famille du défunt de faire face aux problèmes financiers en versant des rentes ou un capital.
- L'assurance Obsèques met l'accent sur les coûts associés aux funérailles, qui sont parfois très élevés.

Pour ces deux types d'assurances décès, des services supplémentaires sont proposés comme la gestion des démarches administratives ou encore un soutien psychologique.

- **Assurance Dépendance**

Dans le domaine de la prévoyance, on retrouve aussi la dépendance, c'est aussi l'un des défis de notre société aujourd'hui. Cela peut être le résultat d'un handicap ou peut être lié à l'âge. Un contrat spécifique peut être signé pour anticiper ce risque et bénéficier du versement de rentes ou de capital, modulables en fonction des cotisations versées (*AG2R LA MONDIALE, 2020*).

- **Arrêt de travail et maintien de salaire**

En raison d'une maladie ou d'un accident, prendre un congé maladie entraîne une baisse de revenu. Le revenu versé par l'assurance maladie obligatoire est en fait de 50% du salaire total.

Certains contrats compensent la perte de revenu pour maintenir le niveau de vie des assurés arrêtés. Selon le contrat souscrit, il est possible de se précautionner contre l'incapacité de travail, la perte totale et irréversible d'autonomie, la maladie de longue durée ou le décès.

- b. **Prévoyance individuelle ou collective**

Les assurés peuvent souscrire individuellement à un ou plusieurs contrats. Cependant, il existe également des plans de prévoyance collective obligatoires ou facultatifs pour les salariés.

L'employeur doit établir un contrat collectif de prévoyance décès pour ses salariés cadres. Sauf si des accords de branche ou des conventions collectives le prévoient, la mise en œuvre de ce dispositif est facultative pour les autres catégories de salariés.

En somme, s'il n'y a quasiment pas de différence entre l'assurance santé et prévoyance en termes d'assurés, l'assurance santé peut être considéré un complément aux frais médicaux ; contrairement à l'assurance prévoyance qui comme une aide financière, soutient l'assuré ou sa famille, au quotidien après un accident de la vie.

Ci-dessous un tableau récapitulatif des garanties en assurance santé et prévoyance

SANTE	PREVOYANCE
<ul style="list-style-type: none"> ✓ Soins courants ✓ Optique ✓ Dentaire ✓ Audition ✓ Pharmacie ✓ Hospitalisation ✓ Médecine douce 	<ul style="list-style-type: none"> ✓ Hospitalisation ✓ Décès / Obsèques ✓ Incapacité ✓ Invalidité

Tableau 1 - Récapitulatif des garanties en Santé et Prévoyance

LA DELEGATION DE GESTION

Aujourd'hui, de nombreuses compagnies d'assurances mandatent totalement ou partiellement leurs activités en confiant le processus de gestion de contrat d'assurance santé et prévoyance à des organismes appelés aussi délégataires. Nous présenterons au travers de ce chapitre les atouts et les limites de ces délégations. Puis, dans la cadre de ces délégations, nous observerons les problématiques auxquelles GENERALI est confronté, pour enfin exposer quelles réponses pouvons-nous apporter avec les solutions Big Data.

I. LA DELEGATION A L'OPPORTUNITE POUR LES ASSUREURS

Comme présenté en introduction, la délégation de gestion des contrats entre l'assureur et le délégataire comprend en général la sous-traitance de la gestion administrative à savoir la création de contrats, la gestion des sinistres, la relation client, etc.

Cette externalisation des activités de gestion des contrats permet aux compagnies d'assurances de se concentrer sur la gestion des risques. De plus, en offrant une relation-client personnalisée par le biais des délégataires, la gestion des contrats au cas par cas rend la collecte d'informations clients plus précises. Cela permet à l'assureur d'effectuer un suivi personnalisé avec ses assurés et de proposer de nouveaux produits plus adaptés (Antenia Solution Assurance, 2018).

II. MISE EN PLACE D'APPLICATIONS DE GESTION DES DELEGATAIRES

a. Création d'applications de suivi de gestion des délégataires

C'est dans ce contexte que nous devons mettre en place, grâce aux solutions Big Data et aux systèmes d'informations de GENERALI, des outils permettant le suivi d'activité de chaque délégataire. A l'aide de ses outils, il sera alors possible :

- De centraliser les informations et avoir une vision 360° des délégataires,
- D'analyser la rentabilité du portefeuille,
- D'étudier la consommation des assurés,
- D'optimiser la gestion du risque,
- De réajuster la tarification en fonction de la sinistralité, etc.

b. Assurer la conformité des activités déléguées

Ces dernières années, le secteur de l'assurance a connu un renforcement des restrictions réglementaires. La RGPD, les nouvelles Directives sur la Distribution d'Assurance (DDA) ainsi que Solvabilité II ont eu un impact fort sur les compagnies d'assurances qui externalisent les activités de gestion. (Antenia Solution Assurance, 2018)

La Directive Solvabilité II est une réforme réglementaire européenne d'assurance qui impose le contrôle et engage la responsabilité des assureurs sur les opérations financières entre le délégataire et celui-ci. Ainsi, les compagnies d'assurances doivent s'assurer à l'aide de reporting et de cartographie des risques, du respect des réglementations.

De même pour la DDA qui, en vigueur depuis le 1^{er} octobre 2018, exerce un contrôle sur la distribution des produits d'assurance. La compagnie d'assurance doit être en mesure de certifier qu'elle s'est conformée à toutes les procédures de vérification imposées par la DDA lors de la distribution du produit.

D'autre part, comme nous l'avons défini précédemment, la RGPD contrôle les données personnelles. Les compagnies d'assurance sont également concernées par cette nouvelle réglementation, qui vise à protéger davantage les personnes qui acceptent de partager leurs données. Étant donné que les entreprises utilisent la délégation, elles devront garantir le respect de l'utilisation des données.

Le respect de ces réglementations représente un véritable chantier pour les assureurs et leurs délégataires. Il est donc nécessaire de réaliser un état des lieux des protocoles considérés comme étant non-conformes afin de mettre en place des actions correctives

III. LES DELEGATAIRES CHEZ GENERALI

L'assureur GENERALI fait partie des nombreuses compagnies d'assurances utilisant la délégation de gestion de contrat santé et prévoyance.

Comme vu précédemment, la délégation permet de proposer aux assurés des produits d'assurance personnalisés, et donc ciblés.

Caractéristique et fonctionnement de la délégation

La signature d'un protocole juridique entre un assureur et un délégataire marque le début de l'externalisation des activités de gestion des contrats. Selon les accords du protocole, le délégataire aura alors pour mission d'émettre les affaires nouvelles souscrites et de gérer le recouvrement des cotisations et prestations. Le délégataire devra en retour faire parvenir à l'assureur, le reporting de son activité comprenant à minima les informations suivantes :

- ✓ le numéro de contrat,
- ✓ la date d'effet de l'adhésion,
- ✓ la dénomination de la garantie souscrite ainsi que les garanties accordées,
- ✓ l'identité et l'adresse de l'assuré,
- ✓ la liste des éventuels bénéficiaires,
- ✓ le montant de la cotisation ainsi que la périodicité de règlement des cotisations.

C'est à partir de ces données que seront construits les suivis des délégataires.

PARTIE III - SOLUTION ANALYTICS - MISE EN PLACE DES SUIVIS

Alain Fernandez définit les solutions Big Data comme étant un « domaine interdisciplinaire qui utilise des techniques d'apprentissage automatique, de la reconnaissance des formes, des statistiques, des bases de données et de la visualisation pour l'extraction d'informations à partir de bases de données volumineuses. » (FERNANDEZ, 2020)

A partir de cette autre définition, nous procéderons à la construction de ces outils en suivant la méthodologie du schéma ci-dessus.

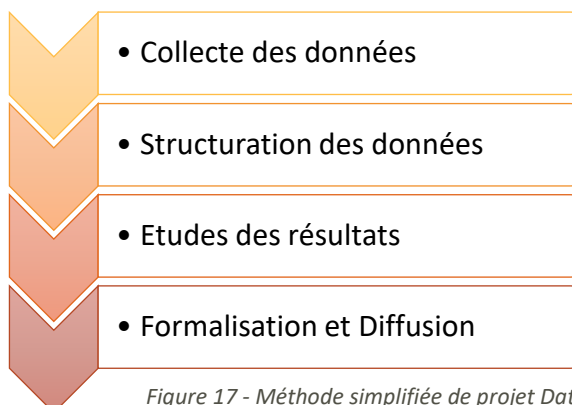


Figure 17 - Méthode simplifiée de projet Data

Nous présenterons la construction des suivis à travers le cas du délégataire n°001. Le même procédé a été réalisé à l'ensemble des délégataires en fonction des données clients disponibles.

I. SUIVI DE RENTABILITE

a. Collecte des données

Un état des lieux a été effectué sur l'ensemble des délégataires afin de détecter l'absence de fichiers. Dans le cas où des fichiers étaient manquants, ceux-ci ont été réclamés auprès des délégataires. Il faut savoir aussi qu'en fonction du protocole d'accord signé entre les deux parties, l'envoi des données peut se faire par mois, par trimestre pour par année. Chaque délégataire fait parvenir son reporting en fonction de ce qui a été convenu dans le protocole et de son fonctionnement. Un travail de structuration, d'uniformisation de la donnée est donc nécessaire.

Pour effectuer un suivi de rentabilité du délégataire, nous avons principalement besoins des données suivantes :

- Primes :
 - Identifiant client
 - Date d'effet du contrat
 - Les garanties souscrites
 - Montant des cotisations toutes taxes comprises (TTC)
 - Montant des cotisations hors taxes (HT)
 - Commission
- Prestations :
 - Identifiant client
 - Date de survenance du sinistre

- Date de règlement du sinistre
- Garantie couvrant le sinistre
- Montant de la prestation

Concernant notre exemple, le délégataire n°001 fait parvenir par mail à GENERALI 2 fois/an un fichier contenant les primes et les commissions, et 1 fois/an le fichier contenant les prestations (Annexe I).

Le premier fichier : *Commissions TNS.xlsx* est composé :

- des données primes et commissions pour les 3 garanties proposées (Décès, Frais médicaux, Incapacité) ;
- des données sinistres pour les garanties Décès et Incapacité

Le deuxième fichier : *Règlement.xlsx* contient l'ensemble des sinistres réglés en santé.

Nous pouvons alors dresser un premier portrait des données à disposition :

Pour des raisons de confidentialité, les données présentées sont des données fictives.

Données PRIME		
Variable	Modalité	Description
NUM_ADH	569	Identifiant client
NOM	DUPONT	Nom
PRENOM	Jacques	Prénom
DATE_EFFET	01/05/2018	Date d'effet
GARANTIE	FM	Garantie souscrite
MT_PRIME_HT	437 €	Montant Prime HT annuelle
COMMISSION	18 €	Commission

Figure 18 - Table Prime du délégataire n°001

Données SINISTRE		
Variable	Modalité	Description
NUM_ADH	569	Identifiant client
DTE_RGL	13/10/2020	Date de règlement du sinistre
DTE_SURV	27/08/2020	Date de survenance du sinistre
GARANTIE	FM	Garantie couvrant le sinistre
MT_PRESTA	14.98€	Montant de la prestation

Figure 19 - Table Sinistre du délégataire n°001

b. Structuration des données

Les fichiers envoyés par le délégataire n°001 sont des fichiers historisés. Nous n'avons pas à concaténer des fichiers de plusieurs années comme nous l'avons fait pour les autres délégataires. Nous devons donc sélectionner la base de données la plus récente. A l'aide de la macro en annexe II, nous avons comparé les dates de création de chaque fichier, pour sélectionner le plus récent.

c. Etudes des résultats

i. La rentabilité

Pour mesurer la rentabilité d'un produit d'assurance, il existe un indicateur de référence chez les assureurs : le Ratio combiné ou S/P.

L'Association des Professionnels de la Réassurance En France la définit comme étant la « *Somme des frais généraux, des commissions encourues, des sinistres survenus et des provisions complémentaires rapportées aux primes acquises* »

Autrement dit, le ratio S/P est l'indicateur qui indique si les primes permettent de couvrir les sinistres.

- Si le S/P est inférieur à 100%, les sinistres sont inférieurs aux primes et l'activité est rentable.
- Si le S/P est supérieur à 100%, les sinistres sont supérieurs aux primes et l'activité est déficitaire.

Cet indicateur a été calculé pour l'ensemble des délégataires par année de survenance, selon la formule :

$$S/P = \frac{\sum \text{des sinistres survenus}}{\sum \text{des cotisations perçues}}$$

ii. Données Délégataire Vs Données GENERALI

GENERALI reçoit aussi dans ses systèmes d'informations les reversements des cotisations et des prestations. Ces données peuvent être extraites à l'aide d'un requêteur (BOBI). Nous allons procéder à la vérification cohérence des données en calculant l'écart entre les primes perçues des délégataires et les données AVTCOLL (système de gestion interne de GENERALI), et les prestations. Cet autre indicateur est calculé selon la formule :

$$\text{Ecart (en \%)} = \frac{\text{Données délégataire} - \text{Données Generali}}{\text{Données délégataire}} * 100$$

d. Formalisation des données

La formalisation ou encore visualisation de données, nous permet d'illustrer l'information. Ainsi, plus les données et informations échangées seront complexes, plus la présentation

devra être éducative. De ce fait, nous avons synthétisé l'information de la rentabilité du délégataire à l'aide d'un tableau.

II. ANALYSE DU PORTEFEUILLE

La table PRIME contient des informations relatives aux données personnelles du client (nom, prénom, date de naissance, adresse). A partir de ces données, nous allons pouvoir analyser la composition du portefeuille de chaque délégataire et déterminer la typologie des assurés propre à chaque délégataire.

a. Structuration des données

Pour notre délégataire n°001, nous avons rencontré plusieurs problèmes de structuration des données. Le format du fichier fournit n'était pas adéquat aux traitements Data. Par exemple, il était difficile pour le programme SAS d'identifier les bénéficiaires d'un contrat. Le fichier comprenait de nombreuses anomalies comme des noms de famille mal orthographiés, etc. Ci-dessous un exemple d'anomalie, où pour un même individu ayant la même date de naissance et prénom : le programme identifie deux personnes distinctes.

NOM	Prénom	Date d'effet	Montant prime 2007
JEAN MARIE	Bernard	01/01/2008	- €
JEAN - MARIE	Bernard	01/01/2008	- €

Figure 20 - Extrait de la base de données Prime de notre délégataire 001

A notre demande, le délégataire nous a fourni un fichier des assurés plus complet et détaillé. Ce qui nous a permis de pallier ces problèmes de fiabilisation des données.

Avec les données géographiques des assurés, nous pouvons obtenir un indicateur sur les régions ou les départements où nous avons le plus d'assurés. A partir de l'adresse de l'assuré, nous avons extrait du code postal le département. Puis, nous avons enrichi ces données à l'aide de cartographie déjà existant dans les bases de GENERALI. Cela nous a permis de rajouter la maille « région » et « département » dans notre base de données.

b. Etudes des résultats

A partir de notre base de données, nous pouvons donc évaluer l'évolution du stock des contrats. Nous disposons des informations suivantes :

- Le nombre total de contrat en cours ;
- Le nombre d'Affaire Nouvelle (AN) ;
- Le nombre de contrat résilié ;
- Le nombre d'AN résilié ;

Nous pouvons alors calculer un autre indicateur : le taux de chute. Cet indicateur est souvent utilisé par les actuaires pour faire un modèle de projection du portefeuille.

c. Formalisation des données

L'analyse du portefeuille permet de dresser le portrait actuel de nos assurés.

A l'aide de plusieurs graphiques de type diagramme circulaire, diagramme en barre ou encore pyramidal, pour notre délégataire n°001, nous observons que la majorité des assurés sont hommes ayant entre 40 et 60 ans. Les assurés souscrivent en moyenne pour deux types de garantie : du Décès et la Santé. Concernant les contrats santé, nombreux sont les assurés ayant inscrit plusieurs bénéficiaires sur un même contrat.

III. ANALYSE DE LA CONSOMMATION

L'analyse de la consommation peut nous éclairer sur les problématiques de rentabilité des certains produits d'assurance. Il permet d'observer la pratique de nos assurés pour les produits d'assurance santé par exemple.

a. Collecte

Les sinistres du délégataire n°001 sont répartis dans deux fichiers : *Commissions.xlsx* pour les garanties décès et revenus de remplacement ; et *Règlement.xlsx* pour la garantie santé. Une base complète de l'ensemble des sinistres a été créée à l'aide d'une jointure entre ces deux bases.

b. Structuration

Un travail de fiabilisation de données a été mené sur les actes médicaux en santé. En effet, il y avait des codes actes de santé qui ne correspondaient à aucun libellé. Après échange avec le délégataire, une cartographie des codes actes a été définis afin de faire correspondre chaque acte médical à une catégorie de soin.

c. Analyse & Visualisation

Dans le suivi de consommation, nous pouvons observer l'âge des assurés qui consomme le plus en santé, on remarque que se sont principalement des personnes âgées.

En termes de fréquence de de consommation (Figure 24), les vacances d'été peuvent expliquer la forte baisse de consommation du mois d'août (fermeture des cabinets médicaux etc.).

Au vu du contexte actuel, il serait intéressant d'analyser les répercussions sanitaire sur le mode de consommation des assurés.

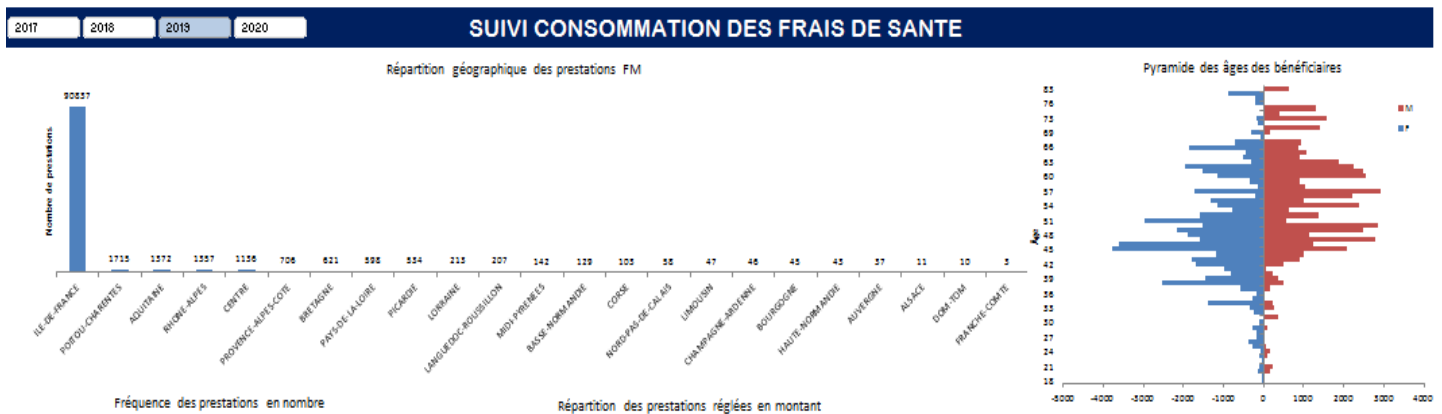


Figure 22 - Suivi de consommation santé du délégataire n°001

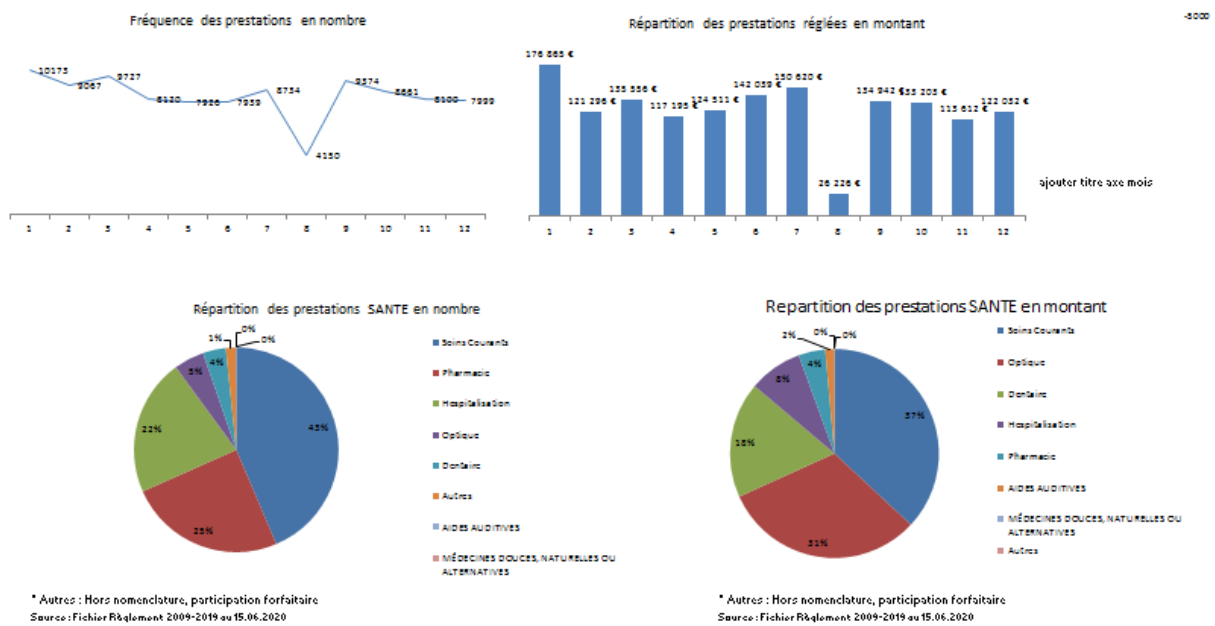


Figure 21- Suivi de consommation santé du délégataire n°001(2)

Ce qui est aussi intéressant d'observer, c'est le rapport entre la fréquence d'un sinistre et son coût. Par exemple, l'optique représente en 2019, 8% des sinistres survenus. Mais il constitue près de 30% des coûts de sinistres. Autrement dit, il y en a très peu, mais ils coûtent cher.

GENERALI peut utiliser ces indicateurs pour revoir la conception d'un produit, proposer une nouvelle tarification si celui-ci n'est pas rentable.

IV. AUTOMATISATION DES SOLUTIONS ANALYTICS

Les suivis des délégataires ont été construits dans des contextes et problématiques différents. Mais les principes de construction sont identiques. Le schéma ci-dessous résume la procédure de construction.



Figure 23 - Principe de construction des suivis des délégataires

Les fichiers Excel envoyés par les délégataires sont centralisés, enrichis à l'aide des données GENERALI. Puis, après traitement du programme SAS (nettoyage, fiabilisation et contrôle), les données alimentent le fichier de suivi d'activité dédié à chaque délégataire. Dans le but de limiter les manipulations et les changements dans le code, et ainsi rendre l'utilisation du programme plus facile pour une personne qui reprendrait ce travail, nous avons essayé de simplifier et d'automatiser l'utilisation des programmes de création de la base de données. Le programme principal est composé de 3 sous programmes :

- Création de la base PRIME
- Création de la base SINISTRE
- Calcul des indicateurs

L'importation des données dans le logiciel est donc automatisée, de même pour le traitement qui prend en compte tout type de format ; et le chargement des données dans le reporting. Des notes de procédures ont été créés afin d'assurer la pérennité des suivis. Ces processus permettent de suivre aussi l'évolution et de mettre à jour les bases de données sans modification de programme.

CONCLUSION

Aujourd'hui, le Big Data cherche à récolter plus en plus de données parfois au détriment de la qualité de celles-ci, les méthodes d'imputation comme celles utilisées dans ce mémoire semble devenir de plus en plus recommandées.

A travers ce mémoire, nous avons pu constater de la complexité de mise en place d'un suivi d'activité : tant au niveau de la collecte, que dans la structuration des données ou l'analyse. L'utilisation de la méthode de gestion de projet « Agile » nous a été bénéfique. En effet, cette méthode nous a permis de rencontrer le délégataire à plusieurs reprises pour renégocier des garanties ou encore les termes du protocole. Ce qui nous a conduit à améliorer notre suivi après chaque rencontre, et d'observer la relation entre un organisme externe et une compagnie d'assurance.

Les différentes méthodologies de traitements de la donnée qu'offre les solutions Big Data, ont été nécessaires pour la construction des suivis d'activités. Ainsi, nous sommes donc en mesure de contrôler à chaque envoi de fichier, la cohérence des informations. Nous pouvons ainsi effectuer un suivi annuel de l'activité de chaque délégataire.

Durant ces deux années d'alternance au sein de l'assureur GENERALI, j'ai éprouvé quelques difficultés dans l'application des cours théoriques sur des cas pratiques comme le traitement de données manquantes ou incohérentes. C'est sans compter sur l'expérience et la pédagogie de ma tutrice, Mme KANZARI ou encore de mes collègues du service Prévoyance Santé que j'ai pu faire face à ces difficultés. Confronté aux différentes problématiques d'une étude statistique, j'ai pu acquérir et d'améliorer mes connaissances statistiques et de programmation ; et développer mon sens de réflexion tout en tenant compte des différentes contraintes qui se présentaient.

En somme, nous avons pu observer les solutions que propose le Big Data dans le domaine de l'Assurance. Cette expérience à GENERALI m'a offert une bonne vision de la statistique au service de l'assurance. Malgré les quelques difficultés rencontrées, cette expérience professionnelle conforte mon désir d'exercer dans le Big Data.

TABLE DES FIGURES

Figure 1 - Les 3V du Big Data	10
Figure 2 - Cycle de vie de la donnée chez GENERALI	11
Figure 3 - Typologie des données utilisées par les assureurs	20
Figure 4 - Les 5 niveaux d'un système d'information []	21
Figure 5 - Les éléments clé de la gouvernance des données	22
Figure 6 - Activité commerciale standard	24
Figure 7 - Activité d'assurance	24
Figure 8 – Principaux pôles d'assurance impactés par le Big Data	25
Figure 9 - Complémentarité des actuaires et des data scientist.....	28
Figure 10 - Méthodologie Cycle en V []	30
Figure 11- Méthode Agile [6]	31
Figure 12 - Implantation de GENERALI dans le monde aujourd'hui	34
Figure 13 - Répartition de l'activité chez GENERALI.....	35
Figure 14 - Organigramme de la Technique Assurance chez GENERALI Assurance	35
Figure 15 - Les différents produits d'assurance chez GENERALI.....	36
Figure 16 - Exemple de remboursement.....	38
Figure 17 - Méthode simplifiée de projet Data.....	45
Figure 18 - Table Prime du délégataire n°001.....	46
Figure 19 - Table Sinistre du délégataire n°001	46
Figure 20 - Extrait de la base de données Prime de notre délégataire 001.....	48
Figure 21 - Suivi de consommation santé du délégataire n°001(2).....	50
Figure 22 - Suivi de consommation santé du délégataire n°001	50
Figure 23 - Principe de construction des suivis des délégataires.....	51

ANNEXE I

SUIVI SP_INVENTAIRE FICHIERS XLS													
Réseau : Déléguaires		Répertoire des fichiers : \\dsi-l4105fio\sas_users_grp333\commun\ps indiv & Prof\006 - Transverse\Delegataire											
Période : 2015-2018													
Système : Av/Call													
Nb :													
Année	Mois	PRIME	SINISTRE	PRIME	SINISTRE	PRIME	SINISTRE	PRIME	SINISTRE	PRIME	SINISTRE	PRIME	SINISTRE
2015	Janvier												
	Février												
	Mars												
	Avril												
	Mai												
	Juin												
	Juillet												
	Août												
	Septembre												
	Octobre												
	Novembre												
	Décembre												
	Janvier												
	Février												
	Mars												
	Avril												
	Mai												
	Juin												

Capture d'écran du fichier de recensement de des données déléguaires

ANNEXE II

```
%macro recherche_table_PTF(day2,month2,year2);

* On identifie la base DSCSANTE_AAAAMMJJ la plus récente au vu de la
date du jour ;
  %LET flag=0;
  %DO %UNTIL(&flag=1);
    %if &month2<10 %then %do;
      %let month2= %sysfunc(cats(0,&month2));
    %end;
    %if &day2<10 %then %do;
      %let day2= %sysfunc(cats(0,&day2));
    %end;
    %let date=&year2&month2&day2;
    %LET S_LOGFILE = "&input./PTF_&date..xlsx";
    %if %sysfunc(fileexist(&S_LOGFILE)) %then %let flag=1;
    %let day2 = %eval(%eval(&day2)-1);
    %if &day2=0 %then %do;
      %let day2=31;
      %let month2 = %eval(%eval(&month2)-1);
    %end;
    %if &month2=0 %then %do;
      %let day2=31;
      %let month2=12;
      %let year2 = %eval(%eval(&year2)-1);
    %end;
  %end;
  %global date_f;
  %let date_f=&date;

*Importation du Fichier récent;
  PROC IMPORT DATAFILE = "&input./PTF_&date_f..xlsx" out= PRIME
    DBMS=XLSX;
  RUN;
%mend recherche_table_PTF;

%recherche_table_PTF(&jj.,&mm.,&aaaa.);
```

Programme de recherche automatique du dernier fichier récent (langage SAS)

BIBLIOGRAPHIE

GARTNER (2020). Inc. and its affiliates. Big Data.

<https://www.gartner.com/en/information-technology/glossary/big-data>

H. ISSARANE (2019). *Les 3 V du Big Data - Volume, Vitesse et Variété*.

<https://le-datascientist.fr/les-3-v-du-big-data>.

A. STROWEL (2018). *Omnia sunt @ommunia : des opera au Big Data*. Revue interdisciplinaire d'études juridiques. Volume 81, n°2, pp. 177-209.

K. STEINBERG (2016). Mémoire de fin d'études « *Qualité des données de santé disponibles en France et de leurs modèles – Comment la garantir pour répondre aux enjeux de la gestion des connaissances médicales ?* ».

D. SRIVASTAVA (2014). *Data quality: the order face of Big Data*.

QUANTMETRY (2020). *La gouvernance des données : par où commencer ?*

<https://www.quantmetry.com/gouvernance-donnees-ou-commencer/>

J.-F. OUTREVILLE (2001). Edito. <http://outreville.com/Edito.htm>

F. LIMOGES (2017). *Données de santé : le big data, outil de lutte contre la fraude*.

<https://www.argusdelassurance.com/acteurs/donnees-de-sante-le-big-data-outil-de-lutte-contre-la-fraude.125092>

PREVISSIMA (2020). *Provisions mathématiques et techniques*.

<https://www.previssima.fr/lexique/provisions-mathematiques.html>

M. KHORDJ, A. HAFFAR, et F. TEULON (2017). *Provisionnement et mesure de risque en assurance dommage dans le cadre de Solvabilité II*. Gestion 2000. vol. 34, n°3, p. 137-168.

G. PIATETSKY (2018). *How many data scientists are there and is there a shortage ?*

KDnuggets. <https://www.kdnuggets.com/how-many-data-scientists-are-there.html/>

Norme NF EN ISO 9000 (2005). *Systèmes de management de la qualité - Principes essentiels et vocabulaire*. Editeur AFNOR. ISBN 5552002200537.

S. RAYNAL, A. L. MEHAUTE, F. ANGOULVANT, et L. FERGUSON (2008). *De la pédagogie de projet à la conception de projet*. La Revue des Sciences de Gestion. n°231-232. p. 53-63.

AZENDOO (2016). *Top 13 des méthodes de gestion de projets*.

<https://blog.azendoo.com/fr/fr/le-top-13-des-methodes-de-gestion-de-projets/>

Q. BRODIER (2020). *Les méthodes de gestion de projet : traditionnelles VS agiles*.

<https://www.attineos.com/blog/autres-sujets/les-methodes-de-gestion-de-projet-traditionnelles-vs-agiles>

GREGORY (2017). *Méthodes de gestion de projet : comment faire son choix ?*
<https://bubbleplan.net/blog/methodes-gestion-projet-comment-choisir/>

CNIL (2018). Le règlement général sur la protection des données - RGPD
<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

CNIL (2018). *Règlement européen sur la protection des données : ce qui change pour les professionnels.* <https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-ce-qui-change-pour-les-professionnels>

CNIL (2018). *Les missions de la CNIL.* <https://www.cnil.fr/fr/les-missions-de-la-cnil>

G. WILHELM (s. d.) *Assurance santé - Définition.* <http://www.assurance-et-mutuelle.com/devis-infos/assurance-sante-definition.html>

N. HERMAL (2013). *Les différents niveaux de couverture santé.* Le Figaro.fr.
<https://www.lefigaro.fr/assurance/2013/07/30/05005-20130730ARTFIG00364-les-differents-niveaux-de-couverture-sante.php>

A. FRUCHARD (2020). *Quels sont les différents niveaux de garantie de la mutuelle santé ?* Réassurez-moi. <https://reassurez-moi.fr/guide/mutuelle-sante/niveau-garanties>

LEGIFRANCE (1989). *Loi EVIN n° 89-1009 du 31 décembre 1989.*
<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000709057/2020-09-28/>

LEGIFRANCE (2019). *Code de la sécurité sociale.*
<https://www.legifrance.gouv.fr/affichCode.do?idSectionTA=LEGISCTA000006172992&cidTexte=LEGITEXT000006073189&dateTexte=vig>

Direction de l'information légale et administrative (2019). *Qu'est-ce que la garantie accidents de la vie ?* <https://www.service-public.fr/particuliers/vosdroits/F3048>

AG2R LA MONDIALE (s. d.) *Assurance prévoyance et dépendance : quelles sont les différences ?* <https://www.ag2rlamondiale.fr/sante-prevoyance/prevoyance/conseil-assurance-prevoyance-et-assurance-dependance-quelle-difference>

ANTENIA (s. d.) *Délégation de gestion : quels enjeux pour les assureurs ? (Partie 1).*
<https://www.antenia.com/2018/07/05/delegation-de-gestion-quels-enjeux-pour-les-assureurs-partie-1/>

ANTENIA (s. d.) *Délégation de gestion : quels enjeux pour les assureurs ? (Partie 2).*
<https://www.antenia.com/2018/08/02/delegation-de-gestion-quels-enjeux-pour-les-assureurs-partie-2/>

A. FERNANDEZ (2020). *Reporting : création et diffusion de rapport d'activité.*
<https://www.piloter.org/business-intelligence/reporting.htm>