

Protection, sauvegarde et archivage
des données médicales informatisées
par les médecins généralistes français
en 2017-2018

THÈSE D'EXERCICE EN MÉDECINE

Présentée à l'Université Claude Bernard Lyon 1

Soutenue publiquement le 09/05/2019

en vue d'obtenir le titre de Docteur en Médecine Générale

par

MAGUY Paul

né le 23/02/1990 à Lyon (69 004)

Sous la direction de

Monsieur LAPRAIS Olivier, Docteur en Médecine Générale

Faculté de Médecine Lyon Est

Liste des enseignants 2018/2019

Professeurs des Universités – Praticiens Hospitaliers

Classe exceptionnelle Echelon 2

BLAY	Jean-Yves	Cancérologie ; radiothérapie
BORSON-CHAZOT	Françoise	Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
COCHAT	Pierre	Pédiatrie
ETIENNE	Jérôme	Bactériologie-virologie ; hygiène hospitalière
GUERIN	Claude	Réanimation ; médecine d'urgence
GUERIN	Jean-François	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
MORNEX	Jean-François	Pneumologie ; addictologie
NIGHOGHOSSIAN	Norbert	Neurologie
NINET	Jean	Chirurgie thoracique et cardiovasculaire
OVIZE	Michel	Physiologie
PONCHON	Thierry	Gastroentérologie ; hépatologie ; addictologie
REVEL	Didier	Radiologie et imagerie médicale
RIVOIRE	Michel	Cancérologie ; radiothérapie
THIVOLET-BEJUI	Françoise	Anatomie et cytologie pathologiques
VANDENESCH	François	Bactériologie-virologie ; hygiène hospitalière

Professeurs des Universités – Praticiens Hospitaliers

Classe exceptionnelle Echelon 1

BOILLOT	Olivier	Chirurgie digestive
BRETON	Pierre	Chirurgie maxillo-faciale et stomatologie
CHASSARD	Dominique	Anesthésiologie-réanimation ; médecine d'urgence
CLARIS	Olivier	Pédiatrie
COLIN	Cyrille	Epidémiologie, économie de la santé et prévention
D'AMATO	Thierry	Psychiatrie d'adultes ; addictologie
DELAHAYE	François	Cardiologie
DENIS	Philippe	Ophtalmologie
DOUEK	Philippe	Radiologie et imagerie médicale
DUCERF	Christian	Chirurgie digestive
DURIEU	Isabelle	Médecine interne ; gériatrie et biologie du vieillissement ; médecine générale ; addictologie
FINET	Gérard	Cardiologie
GAUCHERAND	Pascal	Gynécologie-obstétrique ; gynécologie médicale
GUEYFFIER	François	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
HERZBERG	Guillaume	Chirurgie orthopédique et traumatologique
HONNORAT	Jérôme	Neurologie
LACHAUX	Alain	Pédiatrie
LEHOT	Jean-Jacques	Anesthésiologie-réanimation ; médecine d'urgence
LERMUSIAUX	Patrick	Chirurgie thoracique et cardiovasculaire
LINA	Bruno	Bactériologie-virologie ; hygiène hospitalière
MARTIN	Xavier	Urologie
MERTENS	Patrick	Anatomie
MIOSSEC	Pierre	Immunologie
MOREL	Yves	Biochimie et biologie moléculaire
MORELON	Emmanuel	Néphrologie
MOULIN	Philippe	Nutrition
NEGRIER	Claude	Hématologie ; transfusion
NEGRIER	Sylvie	Cancérologie ; radiothérapie

OBADIA	Jean-François	Chirurgie thoracique et cardiovasculaire
RODE	Gilles	Médecine physique et de réadaptation
TERRA	Jean-Louis	Psychiatrie d'adultes ; addictologie
ZOULIM	Fabien	Gastroentérologie ; hépatologie ; addictologie

Professeurs des Universités – Praticiens Hospitaliers

Première classe

ADER	Florence	Maladies infectieuses ; maladies tropicales
ANDRE-FOUET	Xavier	Cardiologie
ARGAUD	Laurent	Réanimation ; médecine d'urgence
AUBRUN	Frédéric	Anesthésiologie-réanimation ; médecine d'urgence
BADET	Lionel	Urologie
BERTHEZENE	Yves	Radiologie et imagerie médicale
BERTRAND	Yves	Pédiatrie
BESSEREAU	Jean-Louis	Biologie cellulaire
BRAYE	Fabienne	Chirurgie plastique, reconstructrice et esthétique ; Brûlologie
CHARBOTEL	Barbara	Médecine et santé au travail
CHEVALIER	Philippe	Cardiologie
COLOMBEL	Marc	Urologie
COTTIN	Vincent	Pneumologie ; addictologie
COTTON	François	Radiologie et imagerie médicale
DEVOUASSOUX	Mojgan	Anatomie et cytologie pathologiques
DI FILLIPO	Sylvie	Cardiologie
DUBERNARD	Gil	Gynécologie-obstétrique ; gynécologie médicale
DUMONTET	Charles	Hématologie ; transfusion
DUMORTIER	Jérôme	Gastroentérologie ; hépatologie ; addictologie
EDERY	Charles Patrick	Génétique
FAUVEL	Jean-Pierre	Thérapeutique ; médecine d'urgence ; addictologie
FELLAHI	Jean-Luc	Anesthésiologie-réanimation ; médecine d'urgence
FERRY	Tristan	Maladies infectieuses ; maladies tropicales
FOURNERET	Pierre	Pédopsychiatrie ; addictologie
GUENOT	Marc	Neurochirurgie
GUICHAUD	Laurent	Radiologie et imagerie médicale
JACQUIN-COURTOIS	Sophie	Médecine physique et de réadaptation
JAVOUHEY	Etienne	Pédiatrie
JUILLARD	Laurent	Néphrologie
JULLIEN	Denis	Dermato-vénéréologie
KODJIKIAN	Laurent	Ophtalmologie
KROLAK SALMON	Pierre	Médecine interne ; gériatrie et biologie du vieillesse ; médecine générale ; addictologie
LEJEUNE	Hervé	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
MABRUT	Jean-Yves	Chirurgie générale
MERLE	Philippe	Gastroentérologie ; hépatologie ; addictologie
MICHEL	Philippe	Epidémiologie, économie de la santé et prévention
MURE	Pierre-Yves	Chirurgie infantile
NICOLINO	Marc	Pédiatrie
PICOT	Stéphane	Parasitologie et mycologie
PONCET	Gilles	Chirurgie générale
RAVEROT	Gérald	Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
ROSSETTI	Yves	Physiologie
ROUVIERE	Olivier	Radiologie et imagerie médicale
ROY	Pascal	Biostatistiques, informatique médicale et technologies de communication
SAOUD	Mohamed	Psychiatrie d'adultes et addictologie
SCHAEFFER	Laurent	Biologie cellulaire

SCHEIBER	Christian	Biophysique et médecine nucléaire
SCHOTT-PETHELAZ	Anne-Marie	Epidémiologie, économie de la santé et prévention
TILIKETE	Caroline	Physiologie
TRUY	Eric	Oto-rhino-laryngologie
TURJMAN	Francis	Radiologie et imagerie médicale
VANHEMS	Philippe	Epidémiologie, économie de la santé et prévention
VUKUSIC	Sandra	Neurologie

Professeurs des Universités – Praticiens Hospitaliers

Seconde Classe

BACCHETTA	Justine	Pédiatrie
BOUSSEL	Loïc	Radiologie et imagerie médicale
BUZLUCA DARGAUD	Yesim	Hématologie ; transfusion
CALENDER	Alain	Génétique
CHAPURLAT	Roland	Rhumatologie
CHENE	Gautier	Gynécologie-obstétrique ; gynécologie médicale
COLLARDEAU FRACHON	Sophie	Anatomie et cytologie pathologiques
CONFAVREUX	Cyrille	Rhumatologie
CROUZET	Sébastien	Urologie
CUCHERAT	Michel	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
DAVID	Jean-Stéphane	Anesthésiologie-réanimation ; médecine d'urgence
DI ROCCO	Federico	Neurochirurgie
DUBOURG	Laurence	Physiologie
DUCLOS	Antoine	Epidémiologie, économie de la santé et prévention
DUCRAY	François	Neurologie
FANTON	Laurent	Médecine légale
GILLET	Yves	Pédiatrie
GIRARD	Nicolas	Pneumologie
GLEIZAL	Arnaud	Chirurgie maxillo-faciale et stomatologie
GUEBRE-EGZIABHER	Fitsum	Néphrologie
HENAIN	Roland	Chirurgie thoracique et cardiovasculaire
HOT	Arnaud	Médecine interne
HUISSOUD	Cyril	Gynécologie-obstétrique ; gynécologie médicale
JANIER	Marc	Biophysique et médecine nucléaire
JARRAUD	Sophie	Bactériologie-virologie ; hygiène hospitalière
LESURTEL	Mickaël	Chirurgie générale
LEVRERO	Massimo	Gastroentérologie ; hépatologie ; addictologie
LUKASZEWICZ	Anne-Claire	Anesthésiologie-réanimation ; médecine d'urgence
MAUCORT BOULCH	Delphine	Biostatistiques, informatique médicale et technologies de communication
MEWTON	Nathan	Cardiologie
MILLION	Antoine	Chirurgie vasculaire ; médecine vasculaire
MONNEUSE	Olivier	Chirurgie générale
NATAF	Serge	Cytologie et histologie
PERETTI	Noël	Nutrition
POULET	Emmanuel	Psychiatrie d'adultes ; addictologie
RAY-COQUARD	Isabelle	Cancérologie ; radiothérapie
RHEIMS	Sylvain	Neurologie
RICHARD	Jean-Christophe	Réanimation ; médecine d'urgence
RIMMELE	Thomas	Anesthésiologie-réanimation ; médecine d'urgence
ROBERT	Maud	Chirurgie digestive
ROMAN	Sabine	Physiologie
SOUQUET	Jean-Christophe	Gastroentérologie ; hépatologie ; addictologie
THAUNAT	Olivier	Néphrologie
THIBAUT	Hélène	Physiologie
WATTEL	Eric	Hématologie ; transfusion

Professeur des Universités - Médecine Générale

FLORI	Marie
LETRILLIART	Laurent
ZERBIB	Yves

Professeurs associés de Médecine Générale

BERARD	Annick
FARGE	Thierry
LAMBLIN	Gery
LAINÉ	Xavier

Professeurs émérites

BAULIEUX	Jacques	Cardiologie
BEZIAT	Jean-Luc	Chirurgie maxillo-faciale et stomatologie
CHAYVIALLE	Jean-Alain	Gastroentérologie ; hépatologie ; addictologie
CORDIER	Jean-François	Pneumologie ; addictologie
DALIGAND	Liliane	Médecine légale et droit de la santé
DROZ	Jean-Pierre	Cancérologie ; radiothérapie
FLORET	Daniel	Pédiatrie
GHARIB	Claude	Physiologie
GOILLAT	Christian	Chirurgie digestive
MAUGUIERE	François	Neurologie
MELLIER	Georges	Gynécologie
MICHALLET	Mauricette	Hématologie ; transfusion
MOREAU	Alain	Médecine générale
NEIDHARDT	Jean-Pierre	Anatomie
PUGAUT	Michel	Endocrinologie
RUDIGOZ	René-Charles	Gynécologie
SINDOU	Marc	Neurochirurgie
TOURAINE	Jean-Louis	Néphrologie
TREPO	Christian	Gastroentérologie ; hépatologie ; addictologie
TROUILLAS	Jacqueline	Cytologie et histologie

Maîtres de Conférence – Praticiens Hospitaliers Hors classe

BENCHAIB	Mehdi	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
BRINGUIER	Pierre-Paul	Cytologie et histologie
CHALABREYSSE	Lara	Anatomie et cytologie pathologiques
GERMAIN	Michèle	Physiologie
KOLOPP-SARDA	Marie Nathalie	Immunologie
LE BARS	Didier	Biophysique et médecine nucléaire
NORMAND	Jean-Claude	Médecine et santé au travail
PERSAT	Florence	Parasitologie et mycologie
PIATON	Eric	Cytologie et histologie
SAPPEY-MARINIER	Dominique	Biophysique et médecine nucléaire
STREICHENBERGER	Nathalie	Anatomie et cytologie pathologiques
TARDY GUIDOLLET	Véronique	Biochimie et biologie moléculaire

Maîtres de Conférence – Praticiens Hospitaliers

Première classe

BONTEMPS	Laurence	Biophysique et médecine nucléaire
CHARRIERE	Sybil	Nutrition
COZON	Grégoire	Immunologie
ESCURET	Vanessa	Bactériologie-virologie ; hygiène hospitalière
HERVIEU	Valérie	Anatomie et cytologie pathologiques
LESCA	Gaëtan	Génétique
MENOTTI	Jean	Parasitologie et mycologie
MEYRONET	David	Anatomie et cytologie pathologiques
PHAN	Alice	Dermato-vénéréologie
PINA-JOMIR	Géraldine	Biophysique et médecine nucléaire
PLOTTON	Ingrid	Biochimie et biologie moléculaire
RABILLOUD	Muriel	Biostatistiques, informatique médicale et technologies de communication
SCHLUTH-BOLARD	Caroline	Génétique
TRISTAN	Anne	Bactériologie-virologie ; hygiène hospitalière
VASILJEVIC	Alexandre	Anatomie et cytologie pathologiques
VENET	Fabienne	Immunologie
VLAEMINCK-GUILLEM	Virginie	Biochimie et biologie moléculaire

Maîtres de Conférences – Praticiens Hospitaliers

Seconde classe

BOUCHIAT SARABI	Coralie	Bactériologie-virologie ; hygiène hospitalière
BUTIN	Marine	Pédiatrie
CASALEGNO	Jean-Sébastien	Bactériologie-virologie ; hygiène hospitalière
COUR	Martin	Réanimation ; médecine d'urgence
COUTANT	Frédéric	Immunologie
CURIE	Aurore	Pédiatrie
DURUISSEAUX	Michaël	Pneumologie
HAESEBAERT	Julie	Médecin de santé publique
JOSSET	Laurence	Bactériologie-virologie ; hygiène hospitalière
LEMOINE	Sandrine	Physiologie
MARIGNIER	Romain	Neurologie
NGUYEN CHU	Huu Kim An	Pédiatrie Néonatalogie Pharmaco Epidémiologie
ROLLAND	Benjamin	Clinique Pharmacovigilance
SIMONET	Thomas	Psychiatrie d'adultes
		Biologie cellulaire

Maîtres de Conférences associés de Médecine Générale

PIGACHE	Christophe
DE FREMINVILLE	Humbert
ZORZI	Frédéric

Serment d'Hippocrate

Je promets et je jure d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans discrimination.

J'interviendrai pour les protéger si elles sont vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance.

Je donnerai mes soins à l'indigent et je n'exigerai pas un salaire au-dessus de mon travail.

Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement la vie ni ne provoquerai délibérément la mort.

Je préserverai l'indépendance nécessaire et je n'entreprendrai rien qui dépasse mes compétences. Je perfectionnerai mes connaissances pour assurer au mieux ma mission.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé si j'y manque.

Remerciements

Au Professeur LETRILLART, qui me fait l'honneur de présider cette thèse.

Aux Professeurs COLIN et FANTON pour avoir accepté de juger mon travail de thèse malgré les difficultés d'organisation. Au Professeur CHO qui avait aussi accepté, mais n'a pas pu participer pour des raisons d'incompatibilités de plannings indépendantes de ma volonté.

Au Docteur (et co-fondateur de l'Awesome Doctor's News – La clé des infos carabines) Loïc BLANCHET-MAZUEL, qui a accepté au pied levé de faire partie du jury.

Au Docteur LAPRAIS, pour avoir tenu le coup sur la distance malgré mon rythme de travail en dent de scie et mes poussées de stress.

Aux URPS Auvergne Rhône-Alpes, Bourgogne/Franche-Comté, Bretagne, Corse, Grand Est (comprenant Champagne-Ardenne et Lorraine), Guadeloupe, Guyane et Nouvelle-Aquitaine, pour avoir accepté de diffuser le questionnaire sur lequel se base ce travail.

Aux bêta-testeurs du questionnaire, et aux 662 généralistes qui y ont répondu.

Aux Docteurs BUATIER Barbara et MEYER Jean-Hervé, qui m'ont aidé à prendre confiance en moi dans mon travail sur le terrain. Je vous souhaite de tout cœur de continuer à trouver en vous le courage de former mes collègues jeunes généralistes comme vous l'avez fait pour moi, et de pouvoir profiter du fruit de vos années de travail au maximum une fois le moment opportun arrivé.

Aux Docteurs VILLAFANA, AIMARD et FITANIDIS pour leur aide lors de mes premiers pas professionnels en solo.

Au Sieur Jérôme "Mushu/Jay" MICHEL pour sa disponibilité sans faille, son aide indéfectible sur les plans logiciel et matériel, sa bonne humeur, sa camaraderie *IRL* et *online*, et sa capacité à ne pas s'arracher les cheveux pour m'aider à recoller les morceaux quand je fais malencontreusement tout sauter.

Au Docteur en astrophysique Benoît SEIGNOVERT, pour ses conseils avisés notamment concernant le chiffrement SSL avec Letsencrypt, et sa capacité à faire briller encore plus fort qu'elle ne le fait déjà l'une des étoiles les plus lumineuses qui soit.

À monsieur Denis CHENU, du forum d'aide officiel de Limesurvey pour son aide au maniement de Limesurvey (son site : www.sondages.pro).

À Théophile "les statistiques c'est fantastique" TIFFET pour son aide précieuse dans la réalisation de l'analyse statistique de mes résultats.

À mon *bruh* Jules MAGUY, pour son aide sur le CSS de Limesurvey. Et à mon petit frère Augustin MAGUY aussi, parce que je ne l'oublie pas, même s'il n'a pas participé directement à mon travail (issou).

À mes parents, Florence et Stéphane MAGUY, pour avoir conçu, supporté, nourri, logé, blanchi et éduqué aussi bien qu'ils ont pu mes frères et moi.

À mes grands-parents maternels, Marie-Ange et Bruno ANTONIALI, sources de sagesse qui me permettent de garder les pieds sur terre.

À ma grand-mère paternelle, Claire, pour m'avoir laissé grandir les pieds (et le reste) nus dans le gazon. À mon grand-père paternel Jack "*the sky's the limit, Victor Charlie*" MAGUY, que je ne vois que trop peu souvent.

Aux copains de la fac : à Dorian BECAS (et son kouign-amann maison dont je ne me remettrai jamais), à Sidonie la canne et à Ludivine PETIT (la force intellectuelle tranquille) pour les travaux de sous-conf, au Dr "*oh yeah!*" Mariane GUYON, au Dr Paul "Lama" JAULENT, au Dr BARRATIER "désolé, je sors de la douche" Sylvain, Sophie et Antoine GIRARD futurs Docteurs en bio et néphro, au Docteur JAULENT "*I'm not getting any of your shit*" Coline, à l'indestructible Dr ANINO "si-tu-respires-pas, les-orties-ça-pique-pas" Nadine (et à son mangeur de pâtes Clément "Jean Kévin"), et enfin au Dr SANTY Coline (déjà susmentionnée mais avec le nom c'est quand même mieux). Ah, et au Dr Barbara "Barblabla" ABADIE, aussi.

Dans la team CPA, plus particulièrement à Harold "Heula" LECHARPENTIER et à Rachel "la-pharmaco-c'est-rigolo" FONTANA pour m'avoir soutenu pendant notre semestre en commun, puissions-nous rester très bon amis après le travail (Coin !).

Aux supports pendant le parcours : aux membres du Divan (QLPSaV/*Pala tecum* !), à Liliana Osmak-Tizon pour la tasse orange, à la TeamFourStar, à Michael Grzesiek, à Herschel "Guy" Beahm IV (the Doctor), to Daniel Avidan and Dr Brian Wecht (NSP), aux Pentatonix, à Abba, à Franky II et à ma bouilloire.

Au Docteur (mop) Bertille LACHAISE pour m'avoir supporté dans tous les sens du terme pendant plus de neuf ans jusque-là. New-York et le Japon nous attendent !

TABLE DES MATIÈRES

Liste des enseignants 2018/2019	1
Serment d'Hippocrate	6
Remerciements	7
TABLE DES MATIÈRES	9
TABLE DES ILLUSTRATIONS	10
Lexique et abréviations	13
Introduction	15
Recherche bibliographique	16
Travaux de thèse précédents	17
Informatisation en médecine générale	18
En France	18
Un exemple de partage national de données médicales en Grande-Bretagne	20
Type d'installation progicielle des généralistes	21
Installation locale	21
Installation en cloud	22
Définition du risque informatique	24
Distinction entre confidentialité et sécurité	25
Législation informatique et médecine générale	28
Le Règlement Général sur la Protection des Données (RGPD)	28
La certification Hébergeur de Données de Santé (HDS)	29
Problématique	31
Objectifs	31
Matériel et Méthodes	32
Choix de la méthode d'étude	32
Notions d'informatique employées dans le questionnaire	32
Supports du questionnaire informatisé	32
Autorisation par la CNIL via déclaration normale	33
Diffusion du questionnaire	33
Recueil et analyse des données	33
Résultats	35
Réponses au questionnaire	36
Démographie et caractéristiques des répondants	36
Ressenti et connaissances des répondants	37
Perception du risque informatique	38
Pertes de données	38
Méthodes de sauvegarde	40
Restauration de sauvegarde	46
Conduite en cas de panne	47
Archives	47
Score qualité de sauvegarde	48
Discussion	55
Démographie des répondants	55
Ressenti et connaissances des répondants	58
Pertes de données	58

Méthodes de sauvegarde	60
Restauration de sauvegarde	65
Conduite en cas de panne	66
Perception du risque informatique	67
Archives	67
Score selon qualité de sauvegarde	70
Limites de l'étude	71
Points forts de l'étude	73
Réflexions à l'issue du travail	73
Conclusions	77
ANNEXES	79
Fiche de bonnes pratiques	79
Résumé détaillé des travaux de thèse précédents	80
2009 - Dr BRAMI Gregory (PARIS 7)	80
2012 - Dr CHERIET Mondher (PARIS 6)	81
2012 - Dr STAECHÉLÉ Roland (STRASBOURG)	82
Label ASIP e-santé	86
E-mail de sollicitation envoyé via les URPS	88
Récépissé de déclaration simplifiée CNIL	89
Questionnaire imprimable	90
Définitions de notions informatiques	98
Définitions présentes dans le questionnaire	98
Définitions techniques supplémentaires	99
Choix du support logiciel du questionnaire	104
Hébergement physique du questionnaire	105
Hébergement logiciel : installations Limesurvey	105
Installation principale "Rebatel"	106
Concernant la BBox	106
Le Raspberry Pi et Raspbian	107
Installation de Limesurvey sur Raspbian	110
Sauvegarde Limesurvey Rebatel	111
Schéma récapitulatif	114
Incidents techniques "Rebatel"	115
Installation redondante "MiCloud"	115
Différences par rapport à l'installation Rebatel	116
Sauvegarde Limesurvey Micloud	116
Schéma récapitulatif	117
Résultats détaillés	118
Références bibliographiques	120

TABLE DES ILLUSTRATIONS

Figure 1 - Score sur l'intégralité des répondants	49
Figure 2 - Score selon sexe des répondants	50
Figure 3 - Score selon âge des répondants.....	51
Figure 4 - Score selon ancienneté d'installation	52
Figure 5 - Score selon type d'exercice	53
Figure 6 - Âge des répondants.....	55
Figure 7 - Âges des généralistes français en activité en 2018.....	56
Figure 8 – Raspberry Pi 3B.....	107
Figure 9 - Architecture réseau Rebatel.....	114
Figure 10 - Architecture réseau Micloud.....	117
Tableau 1 : Sexe-ratio.....	36
Tableau 2 : Répartition en tranches d'âge.....	36
Tableau 3 : Ancienneté de l'installation	36
Tableau 4 : Mode d'exercice	36
Tableau 5 : Territoire d'exercice.....	36
Tableau 6 : Type d'installation.....	36
Tableau 7 : Partage du progiciel	36
Tableau 8 : Progiciel en <i>cloud</i>	36
Tableau 9 : Intégralité des données de travail stockées dans le progiciel	37
Tableau 10 : Existence d'une sauvegarde (hormis éventuelle sauvegarde par le progiciel en <i>cloud</i>).37	
Tableau 11 : Responsabilité ressentie du médecin généraliste envers ses données s'il ne fait appel à aucun prestataire	37
Tableau 12 : Responsabilité ressentie du médecin généraliste envers ses données s'il en confie la gestion à un prestataire.....	37
Tableau 13 : Connaissance de la machine hébergeant les données de travail	37
Tableau 14 : Connaissance du chemin d'accès informatique des données de travail	37
Tableau 15 : Inclusion dans la sauvegarde des données médicales hors progiciel	37
Tableau 16 : Niveau de confiance dans la sécurité des données de travail.....	38
Tableau 17 : Niveau de compétence personnelle dans la réalisation des sauvegardes	38
Tableau 18 : Niveau de compétence personnelle dans la restauration des sauvegardes.....	38
Tableau 19 : Existence d'une perte de données au cours des 5 dernières années	38
Tableau 20 : Causes retrouvées des pertes de données	39
Tableau 21 : Quantité de données initialement perdues (en jours de travail)	39
Tableau 22 : Présence d'un système de sauvegarde au moment de la perte.....	39
Tableau 23 : Possibilité de restaurer les données perdues à partir d'une sauvegarde	39
Tableau 24 : Moyen de restauration des données.....	39
Tableau 25 : Quantité de données définitivement perdues (en jours).....	40

Tableau 26 : Temps pour retrouver un fonctionnement normal après la perte (en jours).....	40
Tableau 27 : Automatisation des sauvegardes.....	40
Tableau 28 : Personnes réalisant (ou ayant automatisé) la sauvegarde	40
Tableau 29 : Configurations d'intervenants	41
Tableau 30 : Utilisation de l'outil de sauvegarde intégré au progiciel.....	41
Tableau 31 : Méthode de sauvegarde utilisée.....	41
Tableau 32 : Utilisation d'un <i>cloud</i> tiers comme support de sauvegarde	41
Tableau 33 : Exemplaires sur supports locaux.....	42
Tableau 34 : Exemplaires sur supports externalisés	42
Tableau 35 : Exemplaires totaux.....	43
Tableau 36 : Supports locaux	43
Tableau 37 : Supports externalisés	43
Tableau 38 : Supports totaux	44
Tableau 39 : Cas des utilisateurs de progiciel <i>cloud</i> et <i>cloud</i> tiers.....	44
Tableau 40 : Configuration de types de supports (physiques et en <i>cloud</i>).....	45
Tableau 41 : Chiffrement des sauvegardes sur support physique	46
Tableau 42 : Chiffrement des sauvegardes sur <i>cloud</i> tiers.....	46
Tableau 43 : Vérification du processus de sauvegarde	46
Tableau 44 : Test de restauration déjà effectué	46
Tableau 45 : Machines utilisées pour le test de restauration	46
Tableau 46 : Réalisation d'archives informatisées.....	47
Tableau 47 : Durée de conservation d'un dossier informatisé archivé.....	47
Tableau 48 : Supports des archives	47
Tableau 49 : Recopie des archives	47
Tableau 50 : Sexe des répondants	50
Tableau 51 : Âge des répondants	51
Tableau 52 : Score selon ancienneté d'installation.....	52
Tableau 53 : Score selon type d'exercice.....	53
Tableau 54 : Résultats ANOVA.....	54
Tableau 55 : Référentiel fonctionnel label e-santé Logiciel Maisons et Centres de Santé	86
Tableau 56 : Configuration de types de supports (physiques et en <i>cloud</i>) détaillée.....	118
Tableau 57 : Cas des utilisateurs de progiciel <i>cloud</i> et <i>cloud</i> tiers détaillé	119
Tableau 58 : Résultats ANOVA multivariée détaillés.....	119

Lexique et abréviations

ASIP santé : Agence des Systèmes d'Information Partagés de santé. Organisme dépendant du Ministère des Solidarités et de la Santé. Se décrit comme l'Agence Française de la Santé Numérique.

Cloud : Un modèle d'organisation informatique permettant l'accès à des ressources numériques dont le stockage est externalisé sur plusieurs serveurs via Internet.

CNIL : Commission Nationale de l'Informatique et des Libertés. Autorité administrative indépendante française, chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'Homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

CNOM : Conseil National de l'Ordre des Médecins.

DMP : Dossier Médical Partagé (anciennement Dossier Médical Personnel).

DuckDuckGo : Méta-moteur de recherche alternatif, mettant en avant le respect de la vie privée des utilisateurs en ne stockant aucune donnée personnelle, et dont les résultats de recherche ne sont pas filtrés en fonction d'un profil utilisateur.

HDS : Hébergeur de Données de Santé.

MSSanté : Messagerie Sécurisée de Santé. *"Système mis en place par les pouvoirs publics avec l'ensemble des Ordres professionnels afin de [...] permettre à tous les professionnels de santé d'échanger entre eux par email, rapidement et en toute sécurité, des données personnelles de santé de leurs patients, dans le respect de la réglementation en vigueur. Les messageries de l'Espace de Confiance MSSanté possèdent un annuaire commun et certifié de l'ensemble des professionnels de santé."* [\(1\)](#)

NAS : *Network Attached Storage*, en français Stockage Rattaché au Réseau. Le plus souvent, un ou plusieurs disques durs associés dans un boîtier relié à un réseau local, parfois accessibles extérieurement via Internet.

OS : *Operating System*, en français Système d'Exploitation. Les plus populaires sont MacOS chez Apple, Windows chez Microsoft, et les distributions basées sur un noyau Linux telles qu'Ubuntu ou Debian.

Progiciel : Logiciel métier professionnel, ici concernant le médecin généraliste. Les plus connus sont AxiSanté, MédiClick, MédiStory, Weda, Hellodoc, MLM...

RPi : Raspberry Pi ([détails techniques en annexe](#)).

Sécurité informatique : Au sens large, ensemble des moyens techniques, organisationnels, juridiques et humains nécessaires pour conserver, rétablir, et garantir la sécurité des systèmes informatiques. Par extension, permet la protection de données contre une tentative d'accès non autorisée.

Serveur informatique : Un ordinateur dont la fonction est de fournir un service à des ordinateurs clients à distance : accès à un site internet, stockage à distance...

SUDOC : Système Universitaire de Documentation. Catalogue collectif alimenté par l'ensemble des bibliothèques universitaires françaises et de nombreux établissements documentaires de recherche.

TIC : Technologies de l'Information et de la Communication.

VPN : *Virtual Private Network*, en français Réseau Privé Virtuel. Un système permettant de créer un lien direct entre des ordinateurs distants, en isolant ce trafic. ([2](#))

1. Introduction

L'utilisation de l'outil informatique pour stocker les informations médicales est aujourd'hui une nécessité pour tous les médecins. Depuis les années 2000, le taux d'informatisation a bien augmenté : en 2008 près de 90% des généralistes français étaient informatisés, pour atteindre les 96% en 2017 [\(3\)](#).

Une gestion des données médicales adéquate n'est possible que si le médecin en connaît les enjeux : les données médicales sont **sensibles** et **fragiles**. Elles doivent être conservées de façon **pérenne** et **sécurisée**.

Né dans les années 1990, à l'avènement du web, j'ai vécu la démocratisation de l'informatique domestique et l'arrivée de l'Internet haut-débit. Je me suis intéressé de près à ces nouvelles technologies pour mieux les comprendre et les utiliser. J'ai monté moi-même mes propres ordinateurs, utilisé bon nombre de terminaux informatiques, appris les bases de la programmation, et me suis enrichi d'une culture me permettant d'utiliser efficacement cet outil.

Devenu médecin généraliste remplaçant, j'envisage pour mon exercice un environnement de travail équipé des technologies informatiques modernes les plus aidantes pour mon exercice.

Mes stages en médecine générale m'ont amené sur le terrain d'exercice quotidien du généraliste : son cabinet. J'ai pu souvent y remarquer une sécurité informatique souvent négligée, voire délibérément réduite.

Sauter des étapes de sécurisation réduit la complexité des réglages à faire sur une installation informatique. On obtient alors un **système fonctionnel**, mais **vulnérable**.

Un tel système n'est pas à l'abri d'un accident causé par le médecin lui-même, ou d'une personne intervenant sur l'installation. Il peut en résulter une perte des données de santé dont le médecin a la charge.

Une vulnérabilité identifiée peut être révélatrice d'une faiblesse globale du système, et donc d'une exposition à de potentielles attaques malveillantes.

Deux exemples parlants observés sur le terrain :

- un partage réseau complet et non sécurisé de la partition principale du disque dur système de l'ordinateur serveur. Ce mode de partage a permis un partage rapide des données sur le réseau, mais un simple smartphone se raccordant au réseau wifi du groupement permettrait d'accéder aux données, de réaliser un effacement intégral des données du groupement en quelques secondes, puis de rendre inutilisable le serveur.
- une absence de protection physique du serveur contenant l'intégralité des données d'une maison médicale. Situé derrière le comptoir d'accueil, reposant instable sur une étagère, un disque dur externe de sauvegarde branché en suspension sur la prise frontale du boîtier, à proximité immédiate d'un chauffe-eau. Une situation appelant plusieurs accidents possibles, allant de la chute au choc accidentel, au dommage collatéral en cas de dégât des eaux.

L'informatique ne fait pas partie du cœur de métier des médecins, mais est au cœur de leur activité quotidienne. Ces exemples peuvent être la traduction concrète d'une illusion de sécurité apportée par un outil mal connu. Il se pourrait que la méconnaissance - ou le choix délibéré - d'ignorer les problèmes potentiels soient très répandus.

Porter une confiance aveugle dans un système mal connu ou non maîtrisé relèverait alors d'un manque de responsabilité : un défaut qui s'oppose au professionnalisme exigé dans notre métier.

Face à cette constatation, nous avons émis l'hypothèse de travail suivante : les connaissances et pratiques des médecins généralistes en 2017-2018 concernant leur propre installation informatique sont insuffisantes pour garantir la sécurité et la pérennité de leurs données.

1.1. Recherche bibliographique

La bibliographie disponible sur le thème de l'informatique chez le médecin généraliste est peu fournie, et celle portant sur la sauvegarde des données chez le généraliste encore moins. La bibliographie retenue concernant les notions informatiques récentes est bien plus abondante, et se trouve essentiellement sous format numérique.

Devant la rareté de références recherchées dans les banques de données médicales habituelles, les recherches sur le thème de la sécurité informatique se sont donc faites via moteur de recherche type Google et/ou DuckDuckGo. Il n'existe que peu d'ouvrages papier sur ce thème précis.

Concernant les notions informatiques, une recherche Google en anglais a souvent amené davantage de résultats. Les opérateurs booléens (comme des guillemets et opérateurs logiques) ont permis de préciser les recherches.

Les résultats retenus ont autant que possible été sélectionnés pour respecter les critères suivants :

- source d'origine officielle ou fiable
 - Sites d'institutions gouvernementales en ".gouv.fr" : CNIL, ANSSI, HAS, Légifrance...
 - Grands journaux nationaux : Le Figaro, The Guardian
- article exposant ses sources
- information cohérente entre plusieurs sources

Le livre le plus récent portant sur les technologies et méthodes de sauvegarde trouvé est "*Pro Data Backup and Recovery*" de Steven NELSON [\(4\)](#), édité en 2011. Bien que vieux de 7 ans, ce qui peut paraître long en terme d'années d'évolution informatique, il présente les technologies aujourd'hui effectivement en place dans les entreprises et *datacenters* en terme de sauvegarde.

Les mots-clés employés dans le moteur de recherche du catalogue du SUDOC ont été :

médecine générale
informatique
sécurité informatique

protection
perte
données de santé

Les mots-clés employés sur le moteur de recherche de PubMed ont été :

médecine générale / general practice	sauvegarde / backup
informatique / informatics	vol de données / stolen data
sécurité informatique / computer security	brèche de sécurité / data breach
perte de données / data loss	

Les mots-clés employés dans les moteurs de recherche web ont été :

"data loss" 2018 -breach AND statistics	carte vitale histoire	logiciels médicaux médecine générale liste
agrément hébergeur données de santé archival data preservation	CNIL norme simplifiée	lois de confidentialité données de santé
assurance maladie	CNIL RGPD sanctions	loi santé n°2016-41
convention médicale 2016	CNIL sécurité réseau	m disc durability
atlas démographie médicale 2018	CNOM données perte	médecine générale
backup rules	hébergeur responsable	informatique cabinet
big data definition	entreprise data loss cause - breach	ordre des médecins archives durée
big data ethics issues	Google reads emails	RGPD réglementation
biométrie authentification	hard disk data encryption	RGPD sanctions
facteurs multiples	hardware	sécurité informatique guide ANSSI
Cambridge analitica résumé de l'affaire	hébergeur agréé données de santé	SSD data retention time
	healthcare data breach	SSD over provisioning
	INSEE rural urbain définition	USB flash drive reliability
	label logiciel maison de santé	

1.1.1. Travaux de thèse précédents

Via le SUDOC, nous avons dégagé trois précédents travaux de thèse de médecine générale en lien avec la problématique informatique :

- 2009 - Dr BRAMI Gregory (PARIS 7) : Protection des données patients informatisées en médecine générale - Étude transversale - Département du Val d'Oise [\(5\)](#)
- 2012 - Dr CHERIET Mondher - Sécurité des données médicales informatiques en cabinet de médecine générale [\(6\)](#)
- 2012 - Dr STAECHÉLÉ Roland - Sécurité des données informatiques - À propos de la sauvegarde des données de santé au cabinet de médecine générale [\(7\)](#)

Les taux et le nombre de réponses obtenus lors des travaux des thèses précédentes par questionnement indirect ont été faibles : 53% des questionnaires papier pour le Dr BRAMI, 6,93% des questionnaires par e-mail pour le Dr CHERIET.

La problématique de l'informatique n'est pas le cœur de métier du généraliste. Le peu d'intérêt des médecins accordé à leur informatique par rapport au temps disponible rend l'exploration du sujet difficile. La longueur et la complexité des questionnaires des précédents travaux ont été des freins aux réponses.

Les principaux points de faiblesse des méthodes des médecins généralistes évoqués étaient :

- L'existence de praticiens ne réalisant jamais de sauvegardes, et ce encore en 2012,
- une fréquence des sauvegardes insuffisante,
- une insuffisance de vérification de la réussite du processus de sauvegarde,
- une insuffisance des tests de restauration de sauvegarde,
- un taux d'externalisation de sauvegardes insuffisant,
- des méthodes de sauvegardes parfois inadaptées lors de l'utilisation de plusieurs sources (progiciel et données hors progiciel),
 - associées à la méconnaissance de l'emplacement physique et virtuel des données à sauvegarder,
- un sous-emploi de méthodes matérielles de sécurisation répandues, comme le chiffrement logiciel ou matériel,
- une confiance dans la protection du système informatique discordante avec les taux de perte de données,
- une responsabilité ressentie envers la sécurité des données laissées à leur charge insuffisante,
- un taux de pertes de données en phase avec le secteur industriel,
 - associé à un taux de récupération via sauvegarde largement insuffisant.

Un résumé plus détaillé des précédents résultats et conclusions directement en rapport avec la protection et sauvegarde des données est reporté [en annexe](#).

1.2. Informatisation en médecine générale

1.2.1. En France

Il n'existe à ce jour aucune obligation légale d'équipement informatique pour le médecin généraliste exerçant en libéral. Il a cependant obligation de moyens concernant le soin. L'informatique étant un outil de travail au service du soin, on pourrait en déduire que chaque médecin généraliste libéral se doit de s'équiper suffisamment pour améliorer le service rendu. La qualité du soin devrait se trouver améliorée par l'accélération des communications nécessaires au soin, et par la sécurisation des stockages et échanges de données informatisées.

Une des premières incitations à l'équipement informatique via un système national centralisé a été l'apparition de la télétransmission des feuilles de soin électroniques grâce à la carte vitale, diffusée largement en France en 1997 [\(8\)](#). Cependant, ce système peut être autonome et ne nécessite ni progiciel, ni poste informatique pour fonctionner.

La démocratisation de l'accès à Internet et le développement des TIC en général a permis la pénétration du marché des médecins généralistes par des fournisseurs de service informatique. Il y a eu mise en place de différents modes de fonctionnement local :

- sur un seul et même poste pour les médecins travaillant seuls,
- en réseau local pour des médecins associés au sein d'un même bâtiment,
- ou plus tard en réseaux plus larges, incluant les différents soignants au sein d'un ou plusieurs établissements de soin.

Les principaux moteurs d'informatisation sont les recommandations de l'ASIP Santé (Agence des Systèmes d'Information Partagés de Santé). Elles font suite aux évolutions légales des technologies informatiques.

L'impact de ces recommandations et lois se répercute sur les éditeurs de logiciels, qui doivent modifier leurs produits en conséquence. Ainsi, c'est en amont de l'utilisation qu'en font les médecins généralistes qu'interviennent les changements, dans une logique verticale descendante.

Un autre moteur d'informatisation est la Rémunération sur Objectifs de Santé Publique (ROSP). Elle est encadrée par la convention nationale organisant les rapports entre les médecins libéraux et l'Assurance Maladie de 2016. Ses objectifs sont : « *d'aider les médecins :*

- *à investir sur des outils informatiques facilitant le suivi des patients, les échanges avec les autres professionnels et structures intervenant dans la prise en charge des patients,*
- *à développer les démarches d'appui au patient dans le parcours de soins au sein des cabinets médicaux ou en recourant à des services extérieurs. »* [\(10\)](#)

Elle comprend une section dédiée à l'utilisation d'équipements et logiciels informatiques adaptés à l'exercice professionnel. Il s'agit d'une incitation financière, avec rémunération annuelle forfaitaire à la clé. Cette rémunération n'implique pas de retour de données médicales à l'Assurance Maladie.

Les indicateurs du forfait structure de la ROSP sont les objectifs d'informatisation et d'utilisation de télé services aux yeux de l'Assurance Maladie. Ils sont listés comme suit :

- *logiciel métier avec **logiciel d'aide à la prescription** certifié par la HAS, **compatible DMP** pour faciliter le suivi des patients et leur prise en charge coordonnée [...]* ;
- *utilisation d'une **messagerie sécurisée de santé** ;*
- *version du cahier des charges **SESAM-Vitale** intégrant les avenants publiés sur le site du GIE SESAM-Vitale au 31/12 de l'année N-1 par rapport à l'année de référence pour le calcul de la rémunération. [...]*
- *taux de **télétransmission** supérieur ou égal à 2/3, la télétransmission des données fiabilisant la facture et permettant de garantir des délais de paiement très courts ;*
- *affichage des horaires d'ouverture du cabinet dans annuaire santé.* [\(10\)](#)

L'inscription à un service de Messagerie Sécurisée de Santé (MSSanté) type Sisra (initiative régionale en Auvergne-Rhône-Alpes), ou ApiCrypt, est présente pour trois raisons. Premièrement, la plupart des boîtes email classiques peuvent employer un chiffrement SSL des données transmises, mais celui-ci est en option et non pas obligatoire. Deuxièmement, l'hébergement des données sur

une boîte classique ne relève pas obligatoirement d'un agrément d'hébergeur agréé de données de santé, et ne garantit donc pas la non-utilisation des données de santé à des fins détournées. Troisièmement, les emails sont un vecteur d'attaque fréquent, et un client de messagerie sécurisé ajoute une couche de protection supplémentaire. Cependant, il n'y a, à ce jour et à notre connaissance, aucune vérification d'utilisation après inscription à une MSSanté. Les statistiques d'utilisation de messagerie chiffrée reposent donc sur le déclaratif des généralistes. L'adoption de l'envoi des comptes rendus hospitaliers par ce biais se démocratisant, il serait prévisible que l'emploi effectif de ces messageries se généralise.

Pour bénéficier de l'agrément HAS "label e-santé", un logiciel métier doit présenter certaines fonctionnalités de sécurité [\(11\)](#). Les exigences mentionnées dans le document de référence peuvent être qualifiées de basiques (cf. [annexe](#)). Par exemple, aucune mention d'une [authentification à facteurs multiples](#) n'est faite. Les exigences concernant la sauvegarde et l'archivage manquent de précision sur les délais de conservation, ou les méthodes à employer.

1.2.2. Un exemple de partage national de données médicales en Grande-Bretagne

Citons l'exemple de la base de données dédiée à la recherche appelée *UK General Practice Research Database* [\(9\)](#) chez nos confrères britanniques.

Mise en place à l'origine par une compagnie commerciale (VAMP) en 1987, elle a ensuite été rachetée par Reuters et donnée au ministère de la santé britannique. Les médecins généralistes participants se sont vus offrir un ordinateur gratuit en échange d'une saisie des données cliniques standardisées et anonymisées dans la base de données informatisée. Les comptes rendus d'hospitalisation devaient pouvoir être fournis sur demande.

L'objectif était de développer une base de donnée automatisée de recherche clinique, contenant une information longitudinale et continue sur plus de quatre millions de patients, représentative des populations et des pratiques en médecine générale en Grande Bretagne. Cette base de données pouvait être utilisée par l'industrie pharmaceutique pour la surveillance de leurs produits après commercialisation. Elle est désormais gérée par le Bureau des Statistiques Nationales britanniques.

Les médecins reçoivent une modeste contribution financière (à hauteur de 10 pences soit 0,11€ par patient et par an), mais peu considèrent ce paiement comme une véritable incitation. Malgré cela, beaucoup continuent à fournir des données parce qu'ils disent croire en la fourniture d'un service de valeur au Service de Santé National britannique, ou bien parce qu'ils pensent que leur pratique est améliorée par les standards élevés demandés pour le classement des données.

Le financement de l'incitation financière et de l'aide à l'équipement était donc motivé par la récupération des données médicales de terrain sous forme standardisée. Le gain attendu par la compagnie initiatrice était de pouvoir exploiter ou revendre ces données. Après rachat par le

gouvernement, l'utilisation de ces données change et devient le sujet d'analyses statistiques potentiellement utiles en santé publique.

1.3. Progiciel des généralistes

Deux types majoritaires bien distincts d'installations progicielles : une installation locale, ou une installation "en *cloud*". Le marché du progiciel est très concurrentiel. A ce jour, il existe plus d'une trentaine de progiciels destinés aux généralistes, dont moins d'une dizaine fonctionnent en *cloud* [\(10\)](#). Il existe des labels comme ASIP e-santé qui encadrent les fonctions clés des progiciels destinés aux maisons de santé. En dehors de ces labels et du cadre légal, les éditeurs sont libres d'implémenter les fonctions qu'ils souhaitent.

Pour mieux saisir les différences entre les deux types d'installation, résumons les tâches qui incombent au médecin généraliste.

1.3.1. Installation locale

On parle ici d'installation locale au sens où toutes les données médicales circulent dans un système clos, sans export des données vers un service tiers.

Au sens le plus restreint, il s'agit d'un généraliste avec toute son informatique sur un seul ordinateur. Au sens le plus large, il y a partage de données médicales entre différents praticiens de santé partenaires via un réseau local, toujours sans exportation des données vers un tiers non soignant. Les maisons de santé entrent dans cette définition. On inclut également les réseaux plus élargis comme des structures de soin partageant des données entre elles via un [VPN](#) (réseau privé virtuel), car les données restent cloisonnées.

Dans ce cas, le généraliste peut travailler avec le progiciel de son choix, installé sur son poste de travail. Les données du progiciel sont stockées localement, soit sur son poste de travail, soit sur un serveur accessible via le réseau local ou le VPN. C'est idéalement dès la mise en place du progiciel que le généraliste règlera les paramètres de sauvegarde de ses données médicales informatisées.

L'installation initiale effectuée avec l'aide de l'éditeur du progiciel, le généraliste pourra avoir plusieurs tâches d'entretien à effectuer. Si son progiciel fonctionne selon un modèle économique d'abonnement, il faudra le renouveler. Le matériel hébergeant l'installation du progiciel (serveur ou poste de travail du généraliste) devra être entretenu, réparé, mis à niveau ou renouvelé. En cas de fonctionnement en client/serveur, les postes de travail clients devront aussi être mis à jour, idéalement dans le même temps que le serveur, pour limiter les risques d'incompatibilités entre différentes versions du progiciel. Et à chaque mise à jour, le généraliste peut avoir à faire face à des problèmes d'incompatibilité entre son progiciel et les autres logiciels qu'il utilise.

Pour protéger ses données médicales, le généraliste devra sauvegarder la base de données portée par le serveur de son progiciel, idéalement en programmant une automatisation de la tâche.

Pour échanger efficacement et à bon escient les données de patients consultant d'éventuels médecins du même groupement médical, il faudra gérer l'attribution des droits d'accès, avec créations de comptes utilisateurs adaptés à leur fonction dans l'établissement.

Enfin, il faudra sur le long terme renouveler au besoin les périphériques informatiques et consommables : clavier, écran, souris, imprimante et cartouches, scanner...

Le généraliste peut choisir d'être totalement autonome. Ce n'est probablement pas le cas le plus répandu, car l'installation et la gestion autonome d'un système informatique professionnel fiable peut relever d'un métier à part entière, et ce d'autant plus qu'il est étendu et utilisé par de nombreux intervenants.

Autrement, le généraliste peut travailler avec un prestataire de service informatique, ou un tiers plus à l'aise dans ce domaine. Le temps et l'implication nécessaires pour ces opérations varient selon le degré d'implication personnelle du généraliste.

1.3.2. Installation en *cloud*

Le *cloud computing* (ou *informatique en nuage*) a émergé dans les années 2000. Il est défini par le Larousse comme un "*modèle d'organisation informatique permettant l'accès à des ressources numériques dont le stockage est externalisé sur plusieurs serveurs.*" (11). Le mot clé ici est **externalisation** : le temps et les connaissances techniques nécessaires à l'entretien d'un système informatique peuvent être délocalisés chez un tiers, moyennant un abonnement payant.

La démocratisation et l'amélioration des accès à Internet ont permis un essor de ces services, en accélérant la circulation des données entre le fournisseur du service et le client. Le coût de chaque service est réduit par la concentration de moyens chez le fournisseur de service.

Le concept de *cloud computing* est subdivisé en grandes catégories selon la nature du service externe fourni. Celles qui nous intéressent ici sont le *SaaS* (*Software as a Service*, Logiciel en tant que Service) et le *STaaS* (*Storage as a Service*, Stockage en tant que Service) (4).

Dans une installation progicielle "en *cloud*", le généraliste accède via un poste de travail aux données du progiciel sur un serveur distant, situé hors de la structure de soins.

Dans le cas d'un *SaaS* (*Software as a Service*), le médecin généraliste à son cabinet accède, via un poste de travail, au logiciel médical installé sur un serveur distant. Celui-ci stocke ses données sur un ou plusieurs sites hors du cabinet médical.

Selon le progiciel choisi, deux cas de figure se présentent : un fonctionnement avec un progiciel "client" local, ou un système local en "client-léger". Dans le premier cas, l'ordinateur du médecin a son propre système d'exploitation, et le progiciel est une version accédant aux données stockées à distance en temps réel. Il faut alors installer ce progiciel 'client' sur la machine du généraliste. Dans le second cas, l'accès au progiciel est permis via authentification sur une machine distante. Le poste du médecin est alors un "client-léger" ("*thin client*" en anglais, cf. [annexe](#)), son système d'exploitation est minimal, il ne contient aucune des données médicales. Un système d'exploitation sur une machine

distante est le support du progiciel et de ses données. Ainsi, la majeure partie du matériel et l'intégralité du logiciel sont hébergées et gérées par le fournisseur de service.

Toutes les tâches en rapport avec le serveur, dont la maintenance, sont donc délocalisées et gérées par le fournisseur de service. Les frais réglés par le généraliste, dont l'abonnement au progiciel, le sont à un seul interlocuteur.

Le médecin n'a donc besoin localement que de ressources informatiques modestes. Son poste de travail peut être peu coûteux car n'ayant pas ou peu besoin de stockage local (excepté pour son propre système d'exploitation), ni de beaucoup de puissance de calcul car les opérations sont réalisées hors site. Les périphériques d'usage courant (imprimante, écran, clavier, souris) restent nécessaires mais ne représentent qu'un faible coût total.

Il en résulte une plus grande tranquillité d'esprit pour le médecin qui n'a plus à se soucier personnellement de la bonne gestion et sauvegarde de ses données. Toutes les opérations de maintenance se font chez le prestataire de service. S'il a besoin d'intervenir, le prestataire de service n'aura ni à se déplacer ni à recourir à un système d'accès à distance, pas forcément sécurisé. Théoriquement, il en résulte un gain de temps et de sécurité pour les procédures de dépannage.

Un autre avantage des fonctionnements informatiques centralisés est la meilleure circulation des données de santé entre soignants du même réseau. Il est possible d'attribuer des droits d'accès aux données de santé d'un patient uniquement aux personnes concernées par son parcours de soin. Un parfait exemple en est le réseau Sisra de la région Auvergne-Rhône-Alpes [\(12\)](#), où les documents médicaux qui ont résulté des échanges sur le réseau sont intégrables automatiquement dans le progiciel du médecin. Encore une fois, un gain de temps et de sécurité sont attendus.

En contrepartie, il y a **nécessité absolue d'un accès à internet performant et fiable**. Dans l'exemple des "clients-légers", ce type de poste ne peut absolument pas fonctionner s'il n'a pas accès à son serveur. Les données doivent transiter rapidement jusqu'au poste du médecin, c'est-à-dire avec un **débit** suffisant. Le temps entre chaque action du médecin sur son poste et l'exécution par son progiciel doit être minimal, c'est-à-dire avec une **latence** faible. L'usage du service *cloud* ne doit pas impacter les autres utilisations de l'accès Internet en parallèle, que ce soit pour soi-même ou pour les autres utilisateurs du même accès Internet. Les modes d'accès à Internet terrestres répondant le mieux à ces besoins actuellement sont l'accès Internet par fibre optique et l'ADSL 2+. Malheureusement, ils sont encore trop peu accessibles en dehors des grandes villes et du centre des villes de moyenne importance.

L'accès au service en *cloud* via Internet devra être **sécurisé**, donc chiffré. Le mode idéal de connexion est ici via un VPN. A noter que l'emploi de ce type de connexion réduit légèrement le débit maximal entre les ordinateurs, puisqu'il y a entre autre un chiffrement permanent des données échangées. Dans une utilisation bureautique, l'impact en est négligeable.

Idéalement, il faut envisager un accès Internet **redondant** : pour faire face aux pannes des réseaux télécoms terrestres (cuivre ou fibre), il est souhaitable d'avoir une possibilité d'accès à Internet via les

réseaux mobiles. Ceci est permis par l'extension de la couverture du territoire français progressive en 3G, 4G et bientôt 5G (12). Cependant, les modes de connexion inférieurs à la 4G ne sont souhaitables qu'en solution de dépannage, leurs débits et latences pouvant ne pas être suffisants pour assurer un service assez rapide.

Dans le cas d'un *STaaS* (*Storage as a Service*) : le médecin garde une installation progicielle matérielle et logicielle locale. Le généraliste exporte ses sauvegardes et archives hors cabinet, sur un ordinateur externe appartenant au fournisseur de service.

Le médecin confie la gestion de ses sauvegardes et de leurs archives à son prestataire de service, qui se chargera de les chiffrer, réaliser, stocker et dupliquer. Dans ce cas, les gains en matière de sécurité restent théoriquement importants car les opérations sont réalisées par un professionnel.

Les besoins en informatique locale sont par contre plus élevés qu'en *SaaS*, puisqu'un serveur hébergeant le progiciel doit être présent sur place (au cabinet ou groupement), comme dans une installation classique.

Une connexion internet est également indispensable, mais elle peut être moins performante que celle requise pour un *SaaS*. En effet, grâce aux techniques de sauvegarde incrémentale, seule la première sauvegarde est longue, les suivantes ne font qu'envoyer les modifications et/ou ajouts à la base de données. Les envois pourront se faire plutôt hors des horaires de travail, donc n'impactant pas l'utilisation de la connexion Internet en journée.

Ainsi, cette solution de *STaaS* est intermédiaire : elle peut réduire la charge portant sur le médecin, et lui permet de garder la main sur son fonctionnement informatique localement.

1.4. Définition du risque informatique

Pour prendre la mesure des risques pris par des mesures de sécurité limitées, on doit exposer la notion de risque théorique. On se base sur les définitions exposées sur le site eduscol.education.fr (13) :

“Un **risque** est un danger éventuel, plus ou moins prévisible, inhérent à une situation ou à une activité. Le risque est défini juridiquement comme *‘l'éventualité d'un événement futur, incertain ou d'un terme indéterminé, ne dépendant pas exclusivement de la volonté des parties et pouvant causer la perte d'un objet ou tout autre dommage’*.

Les risques sont souvent décrits par type afin d'en faciliter la classification et l'analyse.

De nombreux types de risques existent, ainsi que diverses classifications en catégories et sous-catégories propres à chaque domaine d'activité et / ou profession. [...]

Parmi les différents types de risques, les risques liés à l'information ont des caractéristiques propres. Leurs facteurs sont de plusieurs ordres :

- technologique : dysfonctionnement d'un composant dans une infrastructure technique ou applicative, pouvant perturber la fourniture d'un service, entraîner la perte de confidentialité d'une information ou nuire à l'intégrité du patrimoine informationnel de l'organisation,

- humain : criminalité informatique, intrusion, espionnage industriel, erreurs humaines dans le choix ou l'usage d'une solution informatique,
- risques naturels, essentiellement climatiques : chaleur/froid, inondations..."

Ainsi, les conséquences possibles du risque informatique (14) sont séparables en deux catégories. Celles portant sur la **sécurité** des données sont la perte (ou altération) de données, et la mise hors service d'un système informatique (temporairement ou de façon permanente). Celles portant sur la **confidentialité** des données sont la divulgation d'informations confidentielles, et le déclenchement d'actions pouvant provoquer des accidents physiques ou induire des drames humains.

Concernant les pertes de données : il ne nous a pas été possible de trouver des statistiques récentes et fiables concernant les pertes de données dans le domaine de l'industrie. En effet, les recherches Internet sur le sujet donnent en résultat des sites d'entreprises spécialisées dans la récupération de données perdues. Ces entreprises ont tendance à donner les mêmes statistiques alarmantes, dans un but semble-t-il d'effrayer le lecteur pour l'inciter à employer leurs services (15). Les rares documents disponibles sur ce sujet consistent en des études de marché dont le document PDF récapitulatif est vendu plusieurs milliers de dollars (16), ce qui est évidemment hors de la portée de ce travail de thèse.

Les deux sources les plus fiables reposant sur des méta-analyses alléguées effectuées entre autre par Dell en milieu industriel rapportent comme causes principales des pertes de données, par ordre d'importance : panne matérielle, erreur humaine, corruption de données par logiciel, virus informatique, catastrophes naturelles. Les taux de responsabilité d'erreur humaine et de panne matérielle varient entre 20 et 30%, avec l'erreur humaine souvent citée en première cause (17–20).

1.4.1. Distinction entre confidentialité et sécurité

Nous sommes actuellement à l'heure du "big data". Les multinationales dits "Géants du Web" telles que Google, Amazon, Facebook et Apple en sont les leaders.

Le big data peut être défini par *"Littéralement, [...] mégadonnées, grosses données ou encore données massives. Ils désignent un ensemble très volumineux de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment travailler."* (21). Le big data vise à réaliser l'analyse statistique complexe de quantités astronomiques de données disponibles, et peut porter sur les données personnelles ou de santé. La quantité de données, ainsi que le nombre de machines gérés dans un réseau massif impose un niveau de sécurité maximal, reposant sur les dernières technologies disponibles.

Les hébergements de type *cloud*, proposés par ces mêmes géants du web, sont un support et un outil du big data. Les données qu'ils hébergent font potentiellement partie des données exploitables par le big data. Le service accueillant les données peut théoriquement en faire l'usage qu'il en souhaite, puisqu'il en dispose pleinement sur ses disques durs. Seul un encadrement légal peut

engager la responsabilité du fournisseur de service en cas d'utilisation abusive des données qu'on lui confie.

Les Géants du Web sont leaders en matière de **sécurité** informatique, et ce par nécessité : être exposé publiquement en étant détenteur d'une quantité massive de données peut attiser la convoitise de pirates en tous genres. Cependant, la sécurité apportée par les solutions d'hébergement de données des Géants du Web peut être au détriment de la **confidentialité** des données qu'on leur confie.

Les récents scandales de fuite de données personnelles ont modifié en profondeur le paysage de la confidentialité. Citons l'exemple largement médiatisé du scandale "Cambridge Analytica" datant de 2015 et concernant Facebook, résumé par la CNIL [\(22\)](#) :

"Entre 2013 et 2014 une application tierce développée par des chercheurs pour le compte de Cambridge Analytica a été installée par quelques centaines de milliers d'utilisateurs de Facebook dans le monde (dont 76 Français). Cette application, sous couvert d'un questionnaire de personnalité, demandait également accès aux informations du profil Facebook de l'utilisateur « à des fins de recherche ». Profitant des règles de la « graph API » de Facebook à cette époque (interface mise à disposition des développeurs tiers leur permettant d'interroger la base de données de Facebook des relations entre les utilisateurs), l'application était en réalité alors à même de collecter également des informations personnelles concernant les « amis » de ces utilisateurs. L'ensemble de ces données ont pu être utilisées à des fins de prospection politique. Facebook a annoncé le 4 avril 2018 que 87 millions de personnes étaient concernées dans le monde (dont environ 200 000 Français), 80% étant des citoyens américains."

Dans ce cas, les données personnelles des utilisateurs étaient sécurisées, c'est-à-dire à l'abri d'une destruction ou d'une altération, mais ont été détournées de l'usage présenté à l'utilisateur grâce à un abus des autorisations d'accès aux données personnelles.

Ce scandale a précédé l'entrée en vigueur du [Règlement Général pour la Protection des Données](#), dont nous parlerons plus loin. Ce texte de loi vise à avoir des conséquences concrètes sur la pratique des entreprises et des professionnels libéraux, dont les médecins généralistes.

Jusqu'à l'entrée en vigueur du RGPD, de nombreux services d'hébergement *cloud* gratuits n'étaient pas encadrés par la loi concernant la confidentialité des données. L'adage "Si c'est gratuit, vous êtes le produit." prenait alors son sens. Dans le cas de Google, la traduction visible de l'utilisation des données personnelles est la publicité ciblée proposée à l'utilisateur, de façon parfois insidieuse. Exemple : jusqu'en juin 2017, recevoir un email d'Air France dans sa boîte GMail concernant un vol à venir supposait que ce mail serait lu par un robot. Ces données étaient ensuite injectées dans les algorithmes d'affichage de publicité personnalisée sur les espaces utilisant la technologie AdSense de Google. Ainsi, lors de la navigation sur Internet qui s'ensuivait, des publicités apparaissaient en

proposant des services et objets en rapport avec le lieu d'arrivée du voyage, et ce peu de temps avant le vol en question.

Ce type d'exploitation de données est censé avoir cessé de la part de Google concernant l'aspect personnel, mais continue pour d'autres aspects tels que la sécurité du territoire américain, entre autres. [\(23\)](#)

Le risque concernant l'analyse par les méthodes du big data est corrélé au type de données qui lui sont fournies. On pourrait penser qu'être soumis à de la publicité ciblée en fonction de son activité sur Internet est un risque de perte de confidentialité acceptable dans l'usage d'Internet. Mais ce n'est qu'un aspect visible de l'exploitation des données personnelles, d'autres conséquences possibles sont invisibles et devraient faire réfléchir à chaque numérisation de données personnelles.

Les fuites de données de santé par défaut de **sécurité** existent et sont malheureusement fréquentes [\(24\)](#). Elles concernent cependant plutôt de grandes institutions telles que le MediCare américain, les grands hôpitaux publics nationaux, et sont moins présentes en France car *«il est plus difficile d'obtenir ces données qu'aux États-Unis puisque dans notre pays, le système de santé est éclaté entre plus de 1100 établissements: cliniques de santé privées, cabinets de médecins... Les informations des patients ne communiquent pas ou peu»* [\(25\)](#).

Le problème de **confidentialité** est par contre plus prononcé si l'on considère les données de santé dans le big data [\(26\)](#). On peut imaginer notamment une situation de refus de prêt ou d'assurance chez une personne atteinte d'une maladie grave si les données médicales du patient sont accessibles par l'établissement bancaire au travers des outils du big data. Ces nouvelles technologies ne sont probablement pas bien connues par la population générale. Concernant les médecins généralistes, sont-ils conscients des risques de l'emploi des outils du big data ?

Quoi qu'il en soit, les généralistes doivent respecter le secret médical, donc assurer la confidentialité de leurs données, ce qui implique des mesures de sécurité.

Autant ces risques sont potentiellement méconnus par la majorité des généralistes, autant la gestion confidentielle des données de santé est bien encadrée sur le plan législatif. La loi numéro 2016-41 du 26 Janvier 2016, dans son article 96 I 5° a), a fait entrer dans la loi française un encadrement plus strict des données de santé :

«Toute personne qui héberge des données de santé à caractère personnel recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi social et médico-social, pour le compte de personnes physiques ou morales à l'origine de la production ou du recueil desdites données ou pour le compte du patient lui-même, doit être agréée à cet effet. Cet hébergement, quel qu'en soit le support, papier ou électronique, est réalisé après que la personne prise en charge en a été dûment informée et sauf opposition pour un motif légitime.» [\(28\)](#)

Depuis lors, l'obligation de recourir à un **Hébergeur agréé de Données de Santé** s'impose théoriquement à tous, médecins généralistes compris. L'agrément en question est détaillé plus loin.

1.5. Législation informatique et médecine générale

1.5.1. Le Règlement Général sur la Protection des Données (RGPD)

Le RGPD, règlement européen applicable à chacun des états membres, est entré en vigueur le 25 mai 2018. Plus que jamais, l'accent est mis sur la **protection des données personnelles**. Les médecins généralistes libéraux sont tout autant concernés par ce texte de loi que n'importe quelle entreprise en charge de données personnelles. Jusqu'alors, les généralistes devaient se déclarer à la CNIL en tant qu'hébergeur de données de santé via la Norme Simplifiée n°50, qui a été rendue caduque par l'entrée en vigueur du RGPD [\(27\)](#). Des résumés pragmatiques ont été diffusés aux généralistes par divers biais. Les adhérents de l'URPS Auvergne-Rhône-Alpes ont été informés via la revue "Trait d'Union" numéro 10, parue en octobre 2018 [\(28\)](#).

Dans sa version intégrale, le RGPD détaille dans son article 5, alinéa 1f) les principes à respecter concernant les données personnelles des patients : *"Les données à caractère personnel doivent être [...] traitées de façon à garantir une **sécurité** appropriée des données à caractère personnel, y compris la protection contre le traitement non autorisé ou illicite et contre la perte, la destruction ou les dégâts d'origine accidentelle, à l'aide de mesures techniques ou organisationnelles appropriées (intégrité et **confidentialité**)"*. [\(31\)](#)

Pour les généralistes, cette loi oblige à réaliser un état des lieux du système informatique du médecin, et surtout à consigner le résultat de l'analyse dans un **"registre des activités de traitement"** [\(29\)](#). Nous n'allons pas détailler l'intégralité de la loi, mais nous pouvons surtout noter que le registre des activités de traitement doit contenir :

- à quoi servent [les] données (ce que vous en faites),
- qui accède aux données et à qui elles sont communiquées,
- combien de temps vous les conservez,
- comment elles sont sécurisées.

L'utilité de ce registre est double :

- Forcer un généraliste à réévaluer son installation méthodiquement, donc en repérer les éventuelles failles et le motiver à les corriger,
- servir de base à l'évaluation d'un système informatique par la CNIL en cas de besoin.

En cas de non-respect de l'article 5, c'est-à-dire s'il se produisait une perte, un vol ou une destruction de données personnelles, le médecin est attaquant et est théoriquement passible au maximum de *"jusqu'à 4 % du chiffre d'affaires annuel"* [\(30\)](#).

À ce jour et à notre connaissance, aucune condamnation directe à un généraliste au titre de non-respect du RGPD n'a été portée. C'est par contre le cas pour plusieurs grandes entreprises et hôpitaux. Citons l'entreprise Optical Center [\(31\)](#) et l'hôpital portugais Barreiro-Montijo [\(32\)](#), sanctionnés à hauteur de 250 000 et 400 000€ respectivement.

On peut trouver sur le site de la CNIL (Commission nationale de l'informatique et des libertés) des recommandations [\(33\)](#) aidant à la mise en conformité d'un cabinet médical.

A noter également, une modification intéressante depuis l'entrée en vigueur du RGPD concernant l'accord théorique du patient pour le recueil de données le concernant [\(34\)](#) :

- Vous **devez délivrer aux patients une information portant sur le traitement de données que vous effectuez pour leur prise en charge** (soit dans votre logiciel de suivi, soit dans votre dossier papier). Cela peut être sous la forme d'une affiche, dans votre salle d'attente.
- Vous **n'avez pas besoin de recueillir le consentement des patients pour collecter et conserver les données de santé les concernant**, dans la mesure où leur collecte et leur conservation sont nécessaires aux diagnostics médicaux et à la prise en charge sanitaire ou sociale des patients concernés.

Sa mise en place oblige théoriquement à l'évaluation des systèmes informatiques présents chez les généralistes. La mise en pratique de ces vérifications reste encore théorique, à notre connaissance. En revanche, comme à chaque instauration de nouvelle loi concernant les médecins généralistes avec amende potentielle, une nouvelle escroquerie a fait son apparition. La CNIL a reconnu l'existence de ces pratiques abusives [\(35\)](#) :

“Le RGPD s'applique à toute organisation, publique et privée, qui traite des données personnelles pour son compte ou en tant que sous-traitant, dès lors qu'elle est établie sur le territoire de l'Union européenne ou que son activité cible directement des résidents européens.

Des sociétés profitent de l'entrée en vigueur de ce règlement pour opérer du démarchage auprès des professionnels (entreprises, administrations, associations), parfois de manière agressive, afin de vendre un service d'assistance à la mise en conformité au RGPD.”

L'escroquerie reposait sur la crainte d'une amende au généraliste s'il était en méconnaissance de la loi. Leur existence montre qu'il existait un marché et une rentabilité à les créer, donc qu'il existait de nombreux généralistes mal informés à ce sujet.

1.5.2. La certification Hébergeur de Données de Santé (HDS)

La confiance entre l'utilisateur et son hébergeur des données de santé est une nécessité. Jusqu'à présent, le patient accordait sa confiance au généraliste concernant la conservation de ses données de santé. En plus d'un lien de confiance humain et d'une relation de soin, le cadre légal en est posé par l'obligation du secret médical.

Concernant un hébergement distant, ce lien de confiance humain et de relation de soin n'existe pas. Que ce soit via un progiciel en *cloud* ou simplement un export de sauvegardes chez un hébergeur distant en *cloud*, cette confiance peut être permise grâce à un cadre légal : l'**agrément Hébergeur de Données de Santé**, dans sa version de 2018.

En 2006, l'ASIP santé a créé un "agrément pour l'hébergement de données de santé" (36) qui n'a pas suscité d'engouement lors de sa création, car il "exigeait des conditions relevant d'une forme de protectionnisme, et engendrait des coûts très importants" (37).

La loi du 26 janvier 2016 de modernisation de notre système de santé (38) a prévu le transfert de la procédure d'agrément, jusqu'alors assuré par un comité placé auprès d'Asip santé, à une certification délivrée par des organismes certificateurs agréés.

Depuis le 1er avril 2018, l'ancien agrément a été effectivement remplacé par la certification Hébergeur de Données de Santé (HDS) (39). Cette certification spécifie que *"toute personne physique ou morale qui héberge des données de santé à caractère personnel recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi médico-social pour le compte de personnes physiques ou morales à l'origine de la production ou du recueil de ces données ou pour le compte du patient lui-même, doit être agréée ou certifiée à cet effet."* Cette obligation repose sur l'article L.1111-8 du code de la santé publique, modifié par la loi n° 2016-41 du 26 janvier 2016. (39)

Elle garantit que l'hébergeur est conforme à certaines normes préexistantes, entre autres :

- ISO 20000-1 : Technologies de l'information - Gestion des services, garantir la fiabilité des services informatiques afin d'en permettre sa continuité
- ISO 27001 : Management de la Sécurité de l'Information. Inclut les principes du management qualité de l'ISO 9001, ainsi que le PDCA (*Plan, Do, Check, Act* - Prévoir, Faire, Vérifier, Agir) de l'amélioration continue.
- ISO 27018 : Protection des données à caractère personnel dans l'informatique en nuages

Concrètement, dans le référentiel de certification HDS (40) à sa rubrique 4.4.2, le **détournement d'usage des données médicales** est encadré spécifiquement : *"l'hébergeur ne doit pas utiliser les données de santé qu'il héberge à d'autres fins que l'exécution de la prestation d'hébergement. Est notamment interdite, toute utilisation de ces données à des fins marketings, publicitaires, commerciales, ou statistiques"*.

On retrouve également :

- une **obligation de traçabilité des accès** aux données : *"L'hébergeur doit assurer la journalisation de la transmission des données à caractère personnel à des tiers avec a minima les informations suivantes : la liste des données transmises, le ou les destinataires et les dates de communication."*
- une **obligation d'information au client en cas de sous-traitance** : *"Les clauses contractuelles passées entre l'hébergeur et son client doivent préciser le recours éventuel à un sous-traitant"*

dans le cadre du traitement des données à caractère personnel. Ainsi, l'hébergeur ne doit pas faire appel à un sous-traitant sans l'information préalable du client."

- un **accord de confidentialité et de non-divulgaration** : *"Les contrats de travail des salariés de l'hébergeur doivent inclure une clause de confidentialité. En cas de recours à la sous-traitance, cette exigence s'applique également aux prestataires."*
- un engagement dans une **conservation de la maîtrise des données par le client** : *« L'hébergeur doit avoir défini et formalisé une politique encadrant la mise à disposition et la restitution des données à caractère personnel à ses clients, ainsi que leur destruction. Cette politique doit être communiquée au client sur demande. »*

En sus, des **obligations techniques** sont présentes comme l'obligation de **chiffrement des données avant transfert** via réseau public (dont Internet), l'interdiction d'emploi de supports de stockage portable incompatible avec un chiffrement de données, et une gestion nominative des identifiants et habilitations.

Ces exigences techniques sont précises. Un hébergeur agréé s'astreint à un niveau de sécurité et de confidentialité auquel n'ont jamais été contraints les généralistes, et ce même depuis l'entrée en vigueur du RGPD.

1.6. Problématique

Le fonctionnement en local peut être perçu comme plus sûr par rapport au fonctionnement en *cloud* : contenir les données de santé dans un environnement limité peut sembler rassurant. Il s'agit également de la méthode appliquée par les prestataires de service informatique depuis des années, et donc perçue comme mieux maîtrisée. Elle est fonctionnelle, et peut être perçue comme plus fiable qu'un fonctionnement en *cloud*. On peut penser qu'une installation locale présente moins de points de faiblesse qu'un système en *cloud* car mettant en jeu moins d'acteurs, et que "la force d'une chaîne dépend de son maillon le plus faible".

Étant donné les avantages potentiels d'un système de santé en *cloud* et les évolutions législatives sur ce terrain depuis les derniers travaux de thèse de 2012, on peut s'interroger sur les pratiques actuelles des généralistes. La question posée et à laquelle nous avons tenté de répondre est la suivante :

En 2017-2018, comment les données médicales informatisées sont-elles protégées, sauvegardées et archivées par les médecins généralistes français ?

1.7. Objectifs

Principal : Explorer les pratiques des médecins généralistes libéraux concernant la résilience de leur sauvegarde et archive de données patient, au moyen d'un questionnaire en ligne.

Secondaire : Exposer des mesures simples de consolidation des points perfectibles repérés.

2. Matériel et Méthodes

2.1. Choix de la méthode d'étude

Pour pouvoir comparer nos résultats à ceux des thèses précédentes, il était nécessaire de réaliser comme eux une **étude quantitative**.

Pour interroger directement les généralistes sur leurs pratiques informatiques, l'envoi d'un questionnaire aux médecins était la solution la plus indiquée. Au vu des précédentes thèses sur ce thème, le taux de réponse risquait d'être faible et ce peu importe le moyen d'envoi. Maximiser le nombre de questionnaires envoyés semblait nécessaire. Nous avons donc choisi l'envoi de questionnaires par voie informatique pour une diffusion plus large à moindre frais. Pour maximiser le nombre de répondants potentiels, nous nous sommes astreints à réaliser un questionnaire de durée acceptable en limitant le nombre de problématiques informatiques interrogées.

Pour se concentrer sur les failles potentielles les plus pertinentes à analyser, il a fallu cibler ce qui peut parler concrètement au médecin généraliste : le risque de perte de données. Un des éléments les plus intéressants à évaluer à nouveau en détail était les méthodes de sauvegarde. L'enjeu consistait à trouver le bon positionnement en termes de niveau technique des questions : un niveau technique trop complexe risquait de rebuter les répondants. C'est pourquoi nous avons choisi de concevoir un questionnaire **conditionnel**. Des réponses montrant des connaissances techniques précises sur un thème questionné permettait d'afficher d'autres questions plus poussées. A l'inverse, une réponse négative ou un "Je ne sais pas" permettaient de ne pas questionner en profondeur sur des sujets pas ou mal connus.

2.2. Notions d'informatique employées dans le questionnaire

Pour pouvoir rechercher les insuffisances du système de stockage et de sauvegarde de nos confrères, leur compréhension de certains concepts en informatique est nécessaire. Il nous a semblé nécessaire de présenter des définitions simplifiées sous les intitulés des questions où une notion informatique était requise. Ces précisions ont permis de réduire le risque de biais d'interprétation. Vous trouverez les définitions telles que présentées au répondant au fur et à mesure de ses réponses [en annexe](#).

2.3. Supports du questionnaire informatisé

L'installation matérielle et logicielle a consisté en deux installations redondantes du logiciel Limesurvey, notamment sur Raspberry Pi (RPi) 3B. Le détail de l'installation au domicile du thésard, ainsi que l'installation redondante externalisée, sont détaillées [en annexe](#).

2.4. Autorisation par la CNIL via déclaration normale

Avant l'entrée en vigueur du RGPD, la création d'un questionnaire auto-hébergé nécessitait une déclaration à la CNIL. L'objectif était d'assurer l'emploi de mesures nécessaires et suffisantes à la collecte des données, en restant dans un cadre légal encadré en cas de questionnaires à des patients, par exemple.

Etant donné que le questionnaire contient des réponses apportées par des médecins, et qu'il ne s'agit pas à proprement parler de données médicales, une déclaration simplifiée a été réalisée et un récépissé de déclaration ([en annexe](#)) a été le reçu le 15 mai 2018.

2.5. Diffusion du questionnaire

Les dix-neuf URPS de France et DOM-TOM ont été contactés par mail puis par téléphone afin de proposer la diffusion du questionnaire. Les thèmes abordés ont été présentés via une présentation PowerPoint, et les questions ont été détaillées grâce à une version imprimable du questionnaire où figuraient toutes les questions conditionnelles possibles. En complément, nous avons également sollicité des généralistes dans nos connaissances personnelles, par mail, de la même façon.

Le mail de sollicitation ([en annexe](#)) mettait l'accent sur l'accessibilité du questionnaire, la durée courte prévisible, et l'intérêt même pour des réponses montrant une méconnaissance du sujet. Il comprenait deux adresses internet, chacune pointant vers un des deux exemplaires redondants du questionnaire en ligne.

2.6. Recueil et analyse des données

Le recueil des données a été réalisé automatiquement par le logiciel Limesurvey lors du remplissage des questionnaires par les répondants, pendant une période de 6 mois, du 1/07 au 31/12/2018.

Les données des réponses ont été exportées depuis Limesurvey au format Microsoft Excel. Le logiciel Microsoft Excel 2010 a permis l'analyse statistique des données par la réalisation des tableaux et graphiques, ainsi que de tableaux croisés dynamiques.

Les données récoltées - quantitatives ou semi-quantitatives - ont été analysées en comparant les résultats pertinents des études précédentes avec celles obtenues dans notre questionnaire.

Les résultats ne se prêtant pas à une analyse statistique ont été présentés sous forme de graphique ou tableau, afin d'apporter une analyse descriptive de la situation actuelle.

Afin de rechercher un lien entre les caractéristiques démographiques des répondants et la qualité de leurs méthodes de sauvegarde, nous avons créé un critère composite. Il s'agissait d'un score de qualité de sauvegarde compris entre 0 et 10 points. Les points étaient attribués en fonction des résultats aux questions portant sur les points clé d'une sauvegarde résiliente. Notre score de sauvegarde était une variable quantitative.

Des analyses univariées ont été effectuées afin de déterminer les variables descriptives qui ont une influence statistiquement significative sur le score de sauvegarde.

Nous avons donc employé le test t de Student pour les variables descriptives catégorielles à deux niveaux. Pour les variables descriptives catégorielles à plus de deux niveaux, nous avons employé une ANOVA (analyse de variance) à un facteur.

Des facteurs de confusion étaient suspectés, notamment avec un possible lien entre l'âge et la durée d'installation. On a pris en compte ces possibles biais en réalisant une analyse multivariée sur les variables descriptives dégagées en univarié.

3. Résultats

Sur les dix-neuf URPS français, huit ont accepté de relayer ce questionnaire :

- Auvergne Rhône-Alpes
- Bourgogne/Franche-Comté
- Bretagne
- Corse
- Grand Est (comprenant les URPS Champagne-Ardenne et Lorraine)
- Guadeloupe
- Guyane
- Nouvelle-Aquitaine

Via ces URPS, plus de 8700 mails de sollicitation ont été envoyés à des généralistes du territoire français. Nous n'avons pas de chiffre précis sur le total de mails envoyés, car tous n'ont pas précisé leur total de mails envoyés.

Les refus de diffusion ont été argumentés par :

- Centre/Val de Loire et Pays de la Loire : trop de questionnaires diffusés.
- Hauts de France : n'ont pas souhaité participer car ils n'avaient pas été sollicité lors de la conception du questionnaire.
- Ile de France : pas de réponse malgré de multiples relances à la personne responsable.
- Martinique, Normandie et Océan Indien : ne diffusent que les questionnaires des thésards locaux, car les taux de réponse sont déjà faibles pour les problématiques locales.
- Occitanie : dépasse le cadre des missions fixées localement.
- PACA : ont arrêté de diffuser suite à des faibles taux de réponse aux questionnaires.

Au total, **662 réponses** ont été récoltées (606 sur le serveur "Rebatel", 56 sur le serveur "Micloud"). Parmi elles se trouvaient des réponses partielles que nous avons exclues : 53 au total, 45 sur "Rebatel", 8 sur "Micloud". Nous avons donc retenu 609 réponses complètes (561 + 48 respectivement).

Le temps de réponse moyen au questionnaire complet a été de 6 minutes et 45 secondes : l'objectif d'une durée "entre 3 et 8 minutes" présenté dans le mail de sollicitation a bien été atteint.

Durant la période de recueil de six mois (mi-juillet à fin décembre 2018), deux incidents techniques mineurs se sont produits dans les derniers mois de recueil (détaillés [en annexe](#)), mais n'ont que faiblement impacté le recueil de réponses, au sens où il y a toujours eu au moins un des deux serveurs disponible.

3.1. Réponses au questionnaire

3.1.1. Démographie et caractéristiques des répondants

Tableau 1 : Sexe-ratio

	Effectif	Proportion	Sexe-ratio 0,829
Femmes	276	45,32%	
Hommes	333	54,68%	

Tableau 2 : Répartition en tranches d'âge

Moins de 31 ans	18	2,96%	Médiane "Entre 41 et 50 ans"
Entre 31 et 40 ans	146	23,97%	
Entre 41 et 50 ans	157	25,78%	
51 ans et plus	288	47,29%	

Tableau 3 : Ancienneté de l'installation

Moins de 5 ans	119	19,54%	Médiane "Entre 11 et 20 ans d'installation"
Entre 5 et 10 ans	120	19,70%	
Entre 11 et 20 ans	142	23,32%	
Plus de 21 ans	228	37,44%	

Tableau 4 : Mode d'exercice

Seul	195	32,02%
En groupe	414	67,98%

Tableau 5 : Territoire d'exercice

Urbain	387	63,55%
Rural	222	36,45%

Tableau 6 : Type d'installation

Monoposte	238	39,08%
Client/serveur	371	60,92%

Tableau 7 : Partage du progiciel

Parmi les 414 généralistes exerçant en groupe :

Oui	371	89,61%
Non	41	9,90%
Je ne sais pas	2	0,48%

Tableau 8 : Progiciel en cloud

Parmi les 609 répondants :

Oui	106	17,41%
Non	458	75,21%
Je ne sais pas	45	7,39%

Tableau 9 : Intégralité des données de travail stockées dans le progiciel

Oui	469	77,01%
Non	106	17,41%
Je ne sais pas	34	5,58%

Tableau 10 : Existence d'une sauvegarde (hormis éventuelle sauvegarde par le progiciel en *cloud*)

Oui	527	86,54%
Non	61	10,02%
Je ne sais pas	21	3,45%

3.1.2. Ressenti et connaissances des répondants

Tableau 11 : Responsabilité ressentie du médecin généraliste envers ses données s'il ne fait appel à aucun prestataire

Non	5	0,82%
Oui	579	95,07%
Je ne sais pas	25	4,11%

Tableau 12 : Responsabilité ressentie du médecin généraliste envers ses données s'il en confie la gestion à un prestataire

Non	83	13,63%
Oui	416	68,31%
Je ne sais pas	110	18,06%

Tableau 13 : Connaissance de la machine hébergeant les données de travail

Parmi les 293 répondants avec progiciel classique et fonctionnant en architecture client/serveur :

Oui	273	93,17%
Non	20	6,83%

Tableau 14 : Connaissance du chemin d'accès informatique des données de travail

Parmi les 503 participants n'ayant pas de progiciel en *cloud* :

Oui	280	55,67%
Non	223	44,33%

Tableau 15 : Inclusion dans la sauvegarde des données médicales hors progiciel

Parmi les 117 répondants réalisant une sauvegarde et ne stockant pas l'intégralité de leurs données de travail dans le progiciel :

Oui	60	51,28%
Non	16	13,68%
Je ne sais pas	41	35,04%

3.1.3. Perception du risque informatique

Pour les deux questions suivantes, on a demandé aux participants de coter entre 1 et 5 :

Tableau 16 : Niveau de confiance dans la sécurité des données de travail

1 (faible)	24	3,94%	Moyenne	3,46
2	53	8,70%	Médiane	3
3	229	37,60%	Écart-type	0,96
4	223	36,62%		
5 (fort)	80	13,14%		

74,22% des répondants se positionnent entre 3 et 4, donc plutôt confiants. Seuls 13,14% ont une très forte confiance en la sécurité de leurs données. L'écart-type de 0,96 montre un écart modéré des réponses par rapport à la moyenne de 3,46.

Tableau 17 : Niveau de compétence personnelle dans la réalisation des sauvegardes

Parmi les 527 participants réalisant des sauvegardes :

1 (faible)	70	13,28%	Moyenne	3,28
2	67	12,71%	Médiane	3
3	133	25,24%	Écart-type	1,28
4	158	29,98%		
5 (fort)	99	18,79%		

La majorité des répondants (55,22%) se positionne encore dans les classes 3 et 4, mais la moyenne est plus basse que pour la sécurité des données, et les réponses sont réparties plus uniformément entre les niveaux comme le montre l'écart-type de 1,28.

Tableau 18 : Niveau de compétence personnelle dans la restauration des sauvegardes

Parmi les 527 participants réalisant des sauvegardes :

1 (faible)	181	34,35%	Moyenne	2,50
2	104	19,73%	Médiane	2
3	97	18,41%	Écart-type	1,39
4	88	16,70%		
5 (fort)	57	10,82%		

La médiane est ici à 2, avec donc une majorité des répondants comprise entre les classes 1 et 2 : les répondants sont donc plutôt pessimistes sur leurs compétences de restauration.

3.1.4. Pertes de données

Tableau 19 : Existence d'une perte de données au cours des 5 dernières années

Oui	111	18,23%
Non	498	81,77%

Pour les questions suivantes, les réponses obtenues ne concernent que les 111 participants ayant déclaré une perte de données.

Tableau 20 : Causes retrouvées des pertes de données

Causes isolées	96	86,49%	Causes multiples	15	13,51%
Dysfonctionnement logiciel	34	30,63%	Matérielle + Logicielle	5	1,00%
Panne matérielle	28	25,23%	Humaine + Logicielle	2	0,40%
Erreur humaine	14	12,61%	Humaine + Matérielle	2	0,40%
Virus informatique	11	9,91%	Matérielle + Catastrophe	2	0,40%
Inconnue ou non retrouvée	8	7,21%	Matérielle + Perte	1	0,20%
Perte d'un support	1	0,90%	Humaine + Perte	1	0,20%
Catastrophe naturelle	0	0,00%	Matérielle + Virus	1	0,20%
Vol d'un support	0	0,00%	Logicielle + Perte	1	0,20%

Tableau 21 : Quantité de données initialement perdues (en jours de travail)

Moins d'un mois	92	82,88%	Moyenne	193,70j
Un mois et 6 mois	4	3,60%	Médiane	5j
6 mois et un an	5	4,50%	Écart-type	637,09j
Plus d'un an	10	9,01%		

La quantité moyenne de jours de travail perdus est relativement élevée, du fait des quelques réponses rapportant des durées supérieures à 365 jours. Il y a une grande dispersion des valeurs comme le montre l'écart-type très important.

Tableau 22 : Présence d'un système de sauvegarde au moment de la perte

Oui	106	95,50%
Non	5	4,50%

Tableau 23 : Possibilité de restaurer les données perdues à partir d'une sauvegarde

Parmi les 106 participants ayant perdu des données et disposant d'un système de sauvegarde à ce moment-là :

Oui	55	51,89%	
- la totalité	29	27,36%	52,73%
- une partie	26	24,53%	47,27%
Non	51	48,11%	

Tableau 24 : Moyen de restauration des données

Parmi les 55 participants ayant perdu des données et ayant pu en restaurer l'intégralité :

Sauvegardes locales	0	0,00%
Sauvegardes externalisées	0	0,00%

Ces résultats sont aberrants et probablement dus à un bug de Limesurvey.

Tableau 25 : Quantité de données définitivement perdues (en jours)

Parmi les 26 qui ont pu restaurer des données, mais pas en totalité :

0	3	11,54%	Moyenne	6,35j
<= 1 semaine	18	69,23%	Médiane	2j
> 1 semaine, <= 6 mois	5	19,23%	Écart-type	9,54j
> 6 mois	0	0,00%		

La médiane de 2 jours et l'écart type à 9,54 montrent que les quantités de données perdues varient de plus d'une semaine autour de la moyenne, et que 50% des répondants ont perdu 2 jours ou moins au total.

Tableau 26 : Temps pour retrouver un fonctionnement normal après la perte (en jours)

Parmi les 111 ayant perdu des données :

0 jours	15	13,51%	Moyenne	28,47j
<= 1 semaine	71	63,96%	Médiane	2j
> 1 semaine, <= 6 mois	20	18,02%	Écart-type	123,03j
> 6 mois, <= 1 an	3	2,70%		
> 1 an	2	1,80%		

Même situation que le nombre de jours de perte initiale : l'écart-type est important car un petit nombre de réponses mentionnait une durée longue, cette fois-ci avec des chiffres supérieurs à 180 jours.

3.1.5. Méthodes de sauvegarde

Cette section ne concerne que les 527 participants réalisant des sauvegardes autres que celles éventuellement réalisées au sein d'un *cloud* progiciel. On les appellera par la suite "sauvegardants".

Tableau 27 : Automatisation des sauvegardes

Oui	291	55,22%
Non	231	43,83%
Je ne sais pas	5	0,95%

Concernant les cinq sauvegardants ne sachant pas si leur sauvegarde est automatique : un exerce seul et déclare réaliser lui-même ses sauvegardes via l'outil intégré au système d'exploitation. Les quatre autres sont en groupement, ce sont donc un des autres membres du groupement qui réalisent les sauvegardes.

Tableau 28 : Personnes réalisant (ou ayant automatisé) la sauvegarde

Moi-même	317	60,15%
Autre membre du groupe d'exercice	115	21,82%
Tiers contractuel externe	173	32,83%
Tiers autre	35	6,64%

Tableau 29 : Configurations d'intervenants

Toujours parmi les 527 sauvegardants :

Généraliste seul	229	43,45%
Tiers contractuel seul	131	24,86%
Généraliste + une personne	75	14,23%
Autre membre du cabinet seul	52	9,87%
Tiers autre seul	18	3,42%
Généraliste + 2 personnes	10	1,90%
Généraliste + 3 personnes	3	0,57%
Autres configurations (<1%)	9	1,71%

Nombre d'intervenants		
1	430	81,59%
2	84	15,94%
3	10	1,90%
4	3	0,57%

Le répondant participe-t-il à la sauvegarde ?			Parmi les non participants
Non	210	39,85%	
+ ne connaît pas le plan de continuité	176	33,40%	(83,81%)
+ ne connaît pas le plan de restauration	179	33,97%	(85,24%)
Oui	317	60,15%	

Tableau 30 : Utilisation de l'outil de sauvegarde intégré au progiciel

Oui	322	61,10%
Non	129	24,48%
Je ne sais pas	76	14,42%

Tableau 31 : Méthode de sauvegarde utilisée

Parmi les 129 sauvegardants n'utilisant pas l'outil intégré au progiciel :

Programme tiers	41	31,78%
Copie manuelle	36	27,91%
Outil intégré à l'OS ¹	22	17,05%
Ne sais pas	15	11,63%
Programme tiers + copie manuelle	5	3,88%
Outil intégré à l'OS + copie manuelle	4	3,10%
Outil intégré à l'OS + programme tiers	4	3,10%
Les trois	2	1,55%

Tableau 32 : Utilisation d'un *cloud* tiers comme support de sauvegarde

Oui	56	10,63%	dont progiciel en <i>cloud</i>	7
Non	435	82,54%		
Je ne sais pas	36	6,83%		

- Sauvegardes et supports physiques

On présente ici les réponses à “*combien de sauvegardes avez-vous, par type de support physique ?*”.

Les **supports physiques** sont séparés selon leur lieu de stockage : **locaux** (au cabinet médical) ou **externalisés** (hors cabinet médical). Les champs de réponse se présentaient aux répondants comme suit :

¹[Operating System](#)

Précisez le nombre de sauvegardes par type de **support physique***.

❗ Merci d'indiquer au moins un support de sauvegarde.

	Stocké au cabinet médical	Stocké hors cabinet médical
Clé USB	0	0
Optique = CD/DVD gravé	0	0
Disque dur externe	1	0
↳ dont SSD*	0	0
Disque dur en réseau (NAS)	0	0



Support physique (informatique) : dispositif physique utilisé pour stocker les données informatiques.

Si vous avez plus de 6 supports dans une catégorie, merci d'entrer "6".

*Si vous ne savez pas ce qu'est un **SSD**, laissez à 0.

- Nombre d'exemplaires de sauvegardes

Un seul et même support physique peut contenir plusieurs exemplaires d'une sauvegarde. On parlera ici non pas du nombre de supports physiques, mais du nombre de sauvegardes. Par exemple, un sauvegardant peut conserver deux sauvegardes faites à des dates différentes sur le même support.

Tableau 33 : Exemplaires sur supports locaux

Cloud Locaux	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	97	81	20	3	201	33,00%
1	199	15	14	3	231	37,93%
2	104	0	9	1	114	18,72%
3	29	1	3	0	33	5,42%
4	6	2	1	0	9	1,48%
5	6	0	0	0	6	0,99%
6 et plus	13	0	2	0	15	2,46%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : Moyenne : 1,21 / Médiane : 1 / Écart-type : 1,47

Tableau 34 : Exemplaires sur supports externalisés

Cloud Externalisés	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	154	79	26	5	264	43,35%
1	182	12	20	2	216	35,47%
2	73	6	2	0	81	13,30%
3	29	2	0	0	31	5,09%
4	7	0	0	0	7	1,15%
5 et plus	9	0	1	0	10	1,64%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : Moyenne : 0,93 / Médiane : 1 / Écart-type : 1,20

Tableau 35 : Exemplaies totaux

Cloud Totaux	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	12	70	14	2	98	16,09%
1	118	18	10	4	150	24,63%
2	172	4	15	0	191	31,36%
3	70	4	5	1	80	13,14%
4	40	0	1	0	41	6,73%
5	14	1	2	0	17	2,79%
6	13	1	0	0	14	2,30%
7 et plus	15	1	2	0	18	2,96%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : Moyenne : 2,14 / Médiane : 2 / Écart-type : 2,18

Sur les répondants sans aucun *cloud* : Moyenne : 2,50 / Médiane : 2 / Écart-type : 1,96

Sur les répondants avec au moins un *cloud* : Moyenne : 1,08 / Médiane : 0 / Écart-type : 2,43

On constate que les utilisateurs de *cloud*, progiciel ou tiers, réalisent moins d'exemplaires de sauvegarde que leurs confrères en installation locale.

- Nombre de types distincts de supports physiques utilisés

Tableau 36 : Supports locaux

Cloud Locaux	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	97	81	20	3	201	33,00%
1	248	16	15	4	283	46,47%
2	87	1	14		102	16,75%
3	21				21	3,45%
4	1	1			2	0,33%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : Moyenne : 0,92 / Médiane : 1 / Écart-type : 0,81

Tableau 37 : Supports externalisés

Cloud Externalisés	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	154	79	26	5	264	43,35%
1	233	17	17	2	269	44,17%
2	57	3	6		66	10,84%
3	9				9	1,48%
5	1				1	0,16%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : Moyenne : 0,71 / Médiane : 1 / Écart-type : 0,74

Tableau 38 : Supports totaux

Cloud Totaux	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
0	12	70	14	2	98	16,09%
1	165	20	9	4	198	32,51%
2	183	5	18	1	207	33,99%
3	54	3	5		62	10,18%
4	32		3		35	5,75%
5 et plus	8	1	0	0	9	1,48%
Total général	454	99	49	7	609	100,00%

Sur l'ensemble des répondants : **Moyenne : 1,88 / Médiane : 2 / Écart-type : 1,07**

Sur les répondants sans aucun *cloud* : **Moyenne : 1,91 / Médiane : 2 / Écart-type : 1,08**

Ces trois tableaux montrent bien que les utilisateurs de *cloud* (progiciel ou tiers) utilisent beaucoup moins de supports physiques que leurs confrères en installation locale.

Tableau 39 : Cas des utilisateurs de progiciel *cloud* et *cloud* tiers

Configurations de supports physiques	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
Aucun support physique	70	14	2	86	55,48%
NAS externalisé	10	3	2	15	9,68%
DD externalisé	6	4	1	11	7,10%
Clé USB locale	3	2	1	6	3,87%
Autres combinaisons (fréquence <1%) ²	10	26	1	37	23,87%
Total général	99	49	7	155	100%

Sur les répondants avec au moins un *cloud* : Moyenne : 0,79 / Médiane : 0 / Écart-type : 1,10

Il existe là aussi une multitude de configurations rares de supports (18 en tout). **Sur les 106 utilisateurs de progiciel en *cloud*, la grande majorité (72 soit 67,9%) n'emploie aucun support physique de sauvegarde.** Pour ceux qui doublent leur sauvegarde de progiciel *cloud*, le support préféré est le NAS externalisé, bien que ce soit un cas relativement rare. Ce type de support est moderne et présente l'avantage d'être accessible via le réseau. Il n'y a pas de mobilisation physique du support, donc pas de risque de choc ou de perte.

La proportion de sauvegardants exclusivement sur un *cloud* tiers est réduite : 49 sur 409 soit 8,04%. Parmi eux, 14 soit 28,5% n'emploient aucun support physique.

L'emploi d'un progiciel en *cloud* et d'une sauvegarde sur *cloud* tiers est marginale : 7 soit 1,15% des répondants.

² Pour des raisons de lisibilité, le tableau intégral des combinaisons rares a été reporté [en annexe](#).

Tableau 40 : Configuration de types de supports (physiques et en *cloud*)

On considère ici l'intégralité des répondants. On sépare les répondants utilisateurs de *cloud* progiciel ou tiers, car leur utilisation de supports physiques de sauvegarde est bien inférieure à celle des sauvegardants "non *cloud*".

Cloud Supports physiques	Aucun <i>cloud</i>	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage
DD ³ externe externalisé seul	116	6	4	1	127	20,85%
Aucun support physique	12	70	14	2	98	16,09%
DD externe local et externalisé	71		3	1	75	12,32%
USB externalisé + DD externe local	37	2	2		41	6,73%
USB local seul	34	3	2	1	40	6,57%
USB local et externalisé	27	2	2		31	5,09%
NAS externalisé seul	15	10	3	2	30	4,93%
DD externe local + NAS local	20		4		24	3,94%
DD externe local (dont SSD ⁴ local)	14		4		18	2,96%
USB local et externalisé + DD externe local	11	2			13	2,13%
DD externe local et externalisé + NAS local	8		1		9	1,48%
DD externe local (dont SSD local) + NAS local	8				8	1,31%
DD externe local et externe (dont SSD local et externe)	8				8	1,31%
Autres combinaisons (fréquence <1%) ⁵	75	4	0	8	87	14,29%
Total général	454	99	49	7	609	100%

Il existe une grande variété de combinaisons de supports physiques différentes (49 en tout). La plupart sont peu fréquentes : 87 soit 14,29% des répondants emploient une des 36 combinaisons de supports dont la fréquence d'utilisation est inférieure à 1%.

On dégage ici le nombre de **participants ne réalisant absolument aucune sauvegarde : 12 soit 1,97% des répondants au total.**

Les supports de sauvegarde préférentiels sont par ordre décroissant :

- Installation locale : disque dur externalisé seul, disque dur local et externalisé, USB externalisé et disque dur externe local.
- Progiciel *cloud* seul : aucun support physique, disque dur externalisé, NAS externalisé
- *Cloud* tiers seul : aucun support physique

³ Disque Dur

⁴ *Solid State Drive*, définition [en annexe](#).

⁵ Pour des raisons de lisibilité, le tableau intégral des combinaisons rares a été reporté [en annexe](#).

Tableau 41 : Chiffrement des sauvegardes sur support physique

Parmi les 511 réalisant au moins une sauvegarde sur support physique :

Oui, toutes	83	16,24%
Oui, une partie	26	5,09%
Non	144	28,18%
Je ne sais pas	258	50,49%

Tableau 42 : Chiffrement des sauvegardes sur *cloud* tiers

Parmi les 56 exportant un exemplaire de sauvegarde sur un *cloud* tiers :

Oui	26	46,43%	
- chiffrée avant envoi	15	2,46%	57,69%
- non chiffrée avant envoi	1	0,16%	3,85%
- ne savent pas quand est chiffrée	10	1,64%	38,46%
Non	5	8,93%	
Je ne sais pas	25	44,64%	

Tableau 43 : Vérification du processus de sauvegarde

Parmi les 491 n'utilisant pas de progiciel *cloud* et réalisant une sauvegarde :

Oui	280	57,03%	
- Journal de sauvegarde seul	93	15,27%	33,21%
- Fichiers sauvegarde seuls	102	16,75%	36,43%
- Journal + fichiers	38	6,24%	13,57%
- Non précisé	47	7,72%	16,79%
Non	211	42,97%	

3.1.6. Restauration de sauvegarde

Parmi les 527 participants réalisant des sauvegardes :

Tableau 44 : Test de restauration déjà effectué

Oui	135	25,62%
- Une seule fois en tout	37	
- Moins d'une fois par an	45	
- Une fois par an ou plus souvent	53	
Non	392	74,38%

Tableau 45 : Machines utilisées pour le test de restauration

Hors base de production	89	65,93%
Sur base de production	44	32,59%
Les deux	2	1,48%

3.1.7. Conduite en cas de panne

Sur les 609 répondants :

- 179 (29,39%) connaissaient le plan de continuité à suivre en cas de panne,
- 194 (31,86%) connaissaient le plan de restauration.

3.1.8. Archives

Tableau 46 : Réalisation d'archives informatisées

Oui	157	25,78%
Non	452	74,22%

Tableau 47 : Durée de conservation d'un dossier informatisé archivé

Parmi les 157 archiveurs :

<= 10 ans	51	32,48%	Moyenne	23 ans
> 10 et < 20 ans	8	5,10%	Médiane	20 ans
> 20 et <= 30 ans	87	55,41%	Écart-type	19,47 ans
> 30 ans	11	7,01%		

Plus de la moitié des répondants estime correctement la durée réglementaire d'archivage conseillée par l'ordre des médecins.

Tableau 48 : Supports des archives

Parmi les 157 archiveurs :

Configuration de supports			Supports par archiveur		
DD ⁶ externe seul	91 (12)	57,96%	1	124	78,98%
Clé USB + DD externe	18 (2)	11,46%	2	31	19,75%
Cloud seul	12	7,64%	3	1	0,64%
Clé USB seule	12	7,64%	4	1	0,64%
DD externe + Cloud	8 (3)	5,10%			
Optique standard seul	6	3,82%	Lieu de stockage		
M-Disc seul	3	1,91%	Au cabinet seul	42	26,75%
DD externe + Optique standard	2	1,27%	Hors cabinet seul	82	52,23%
Configurations rares (<1%)	5	3,18%	Les deux	31	19,75%

Les SSD sont parmi les disques durs, et figurent entre parenthèses.

Tableau 49 : Recopie des archives

Oui	34	21,66%		
- Tous les ans	25		73,53%	
- Tous les 2 ans	3		2,44%	
- > tous les 2 ans	6		1,33%	
Non	123	78,34%		

⁶ Disque dur

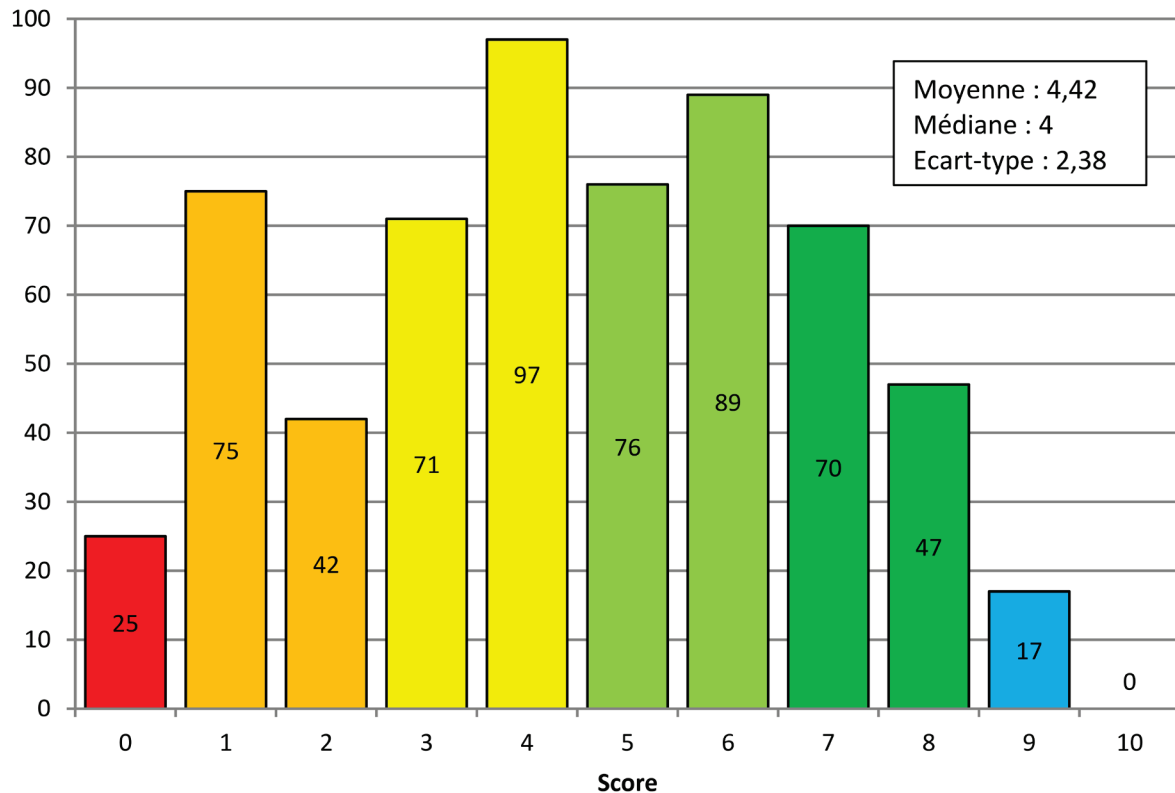
3.2. Score qualité de sauvegarde

Afin de rechercher un lien entre les caractéristiques démographiques des répondants et la qualité de leurs méthodes de sauvegarde, nous avons créé un critère composite. Il s'agit d'un **score de qualité de sauvegarde** compris entre 0 et 10 points.

Les points étaient attribués en fonction des résultats aux questions portant sur les points clés d'une sauvegarde résiliente :

- Intégralité des données médicales sauvegardées
 - Intégralité des données dans le progiciel OU données médicales hors progiciel incluses dans les sauvegardes : 1 point
- Respect des éléments de la [règle des 3-2-1](#)
 - Trois exemplaires des données de travail : 1 point
 - Sur deux types de supports différents : 1 point
 - Dont une copie hors site : 1 point
- Sauvegardes automatisées : 1 point
- Réussite du processus de sauvegarde vérifiée : 1 point
- Utilisation de l'outil de sauvegarde intégré au progiciel : 1 point
- Sauvegardes chiffrées : 1 point
- Test de restauration : selon la fréquence
 - Si effectué une fois seulement : 1 point
 - Si effectué plus d'une fois : 2 points

Figure 1 - Score sur l'intégralité des répondants



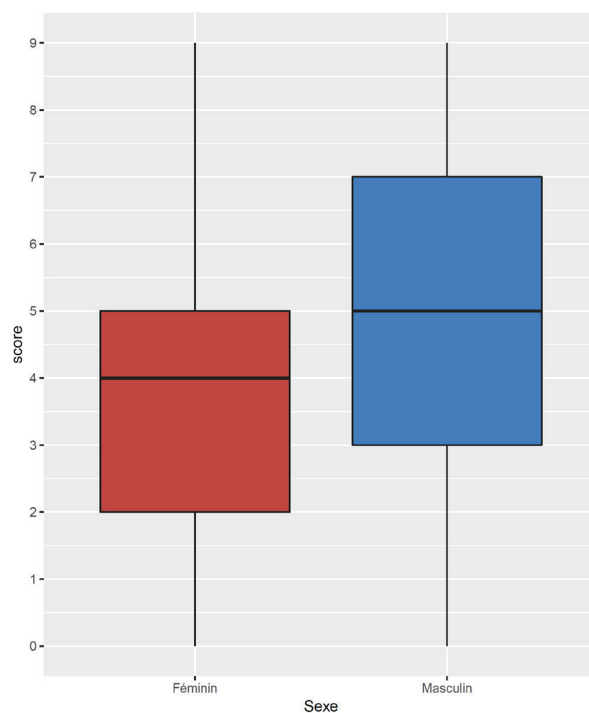
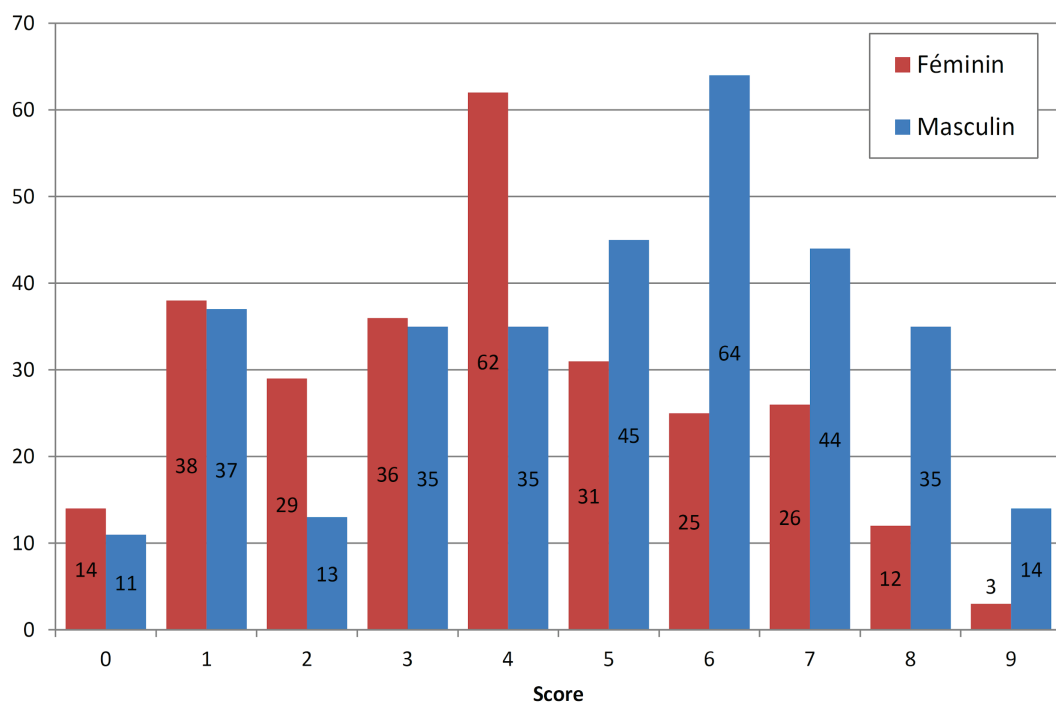
Globalement, une **moyenne à 4,42/10** ainsi que l'**absence de répondant avec un score supérieur à 9/10** nous montrent que les pratiques de sauvegarde de nos répondants sont perfectibles. La répartition des scores reste grossièrement gaussienne centrée autour de la médiane, sans majorité clairement définie.

Nous avons recherché si certaines des caractéristiques démographiques de nos répondants (les facteurs d'étude) étaient en lien avec les scores de sauvegarde (les variables de réponse).

Nous avons cherché à réaliser une analyse multivariée : elle a tenu compte de liens potentiels entre nos facteurs d'études (notamment âge ancienneté d'installation), et des potentiels facteurs de confusion.

Il faut au préalable réaliser des analyses univariées pour dégager les facteurs d'études en lien avec notre score. Dans le cas d'un facteur d'études à deux groupes, nous avons réalisé des tests t de Student. Dans le cas des facteurs d'étude qualitatifs à plus de deux groupes, nous avons réalisé des analyses de variance (ANOVA) à un facteur.

Figure 2 - Score selon sexe des répondants



Les boîtes représentent la répartition en quartiles des scores.

Tableau 50 : Sexe des répondants

	Féminin	Masculin
Moyenne	3,85	4,90
1 ^{er} quartile	2	3
Médiane	4	5
3 ^{ème} quartile	5	7
Écart-type	2,21	2,41

Test t de Student
p=3,321E-08

Parmi nos répondants, les hommes ont de meilleurs scores de sauvegarde que les femmes.

Figure 3 - Score selon âge des répondants

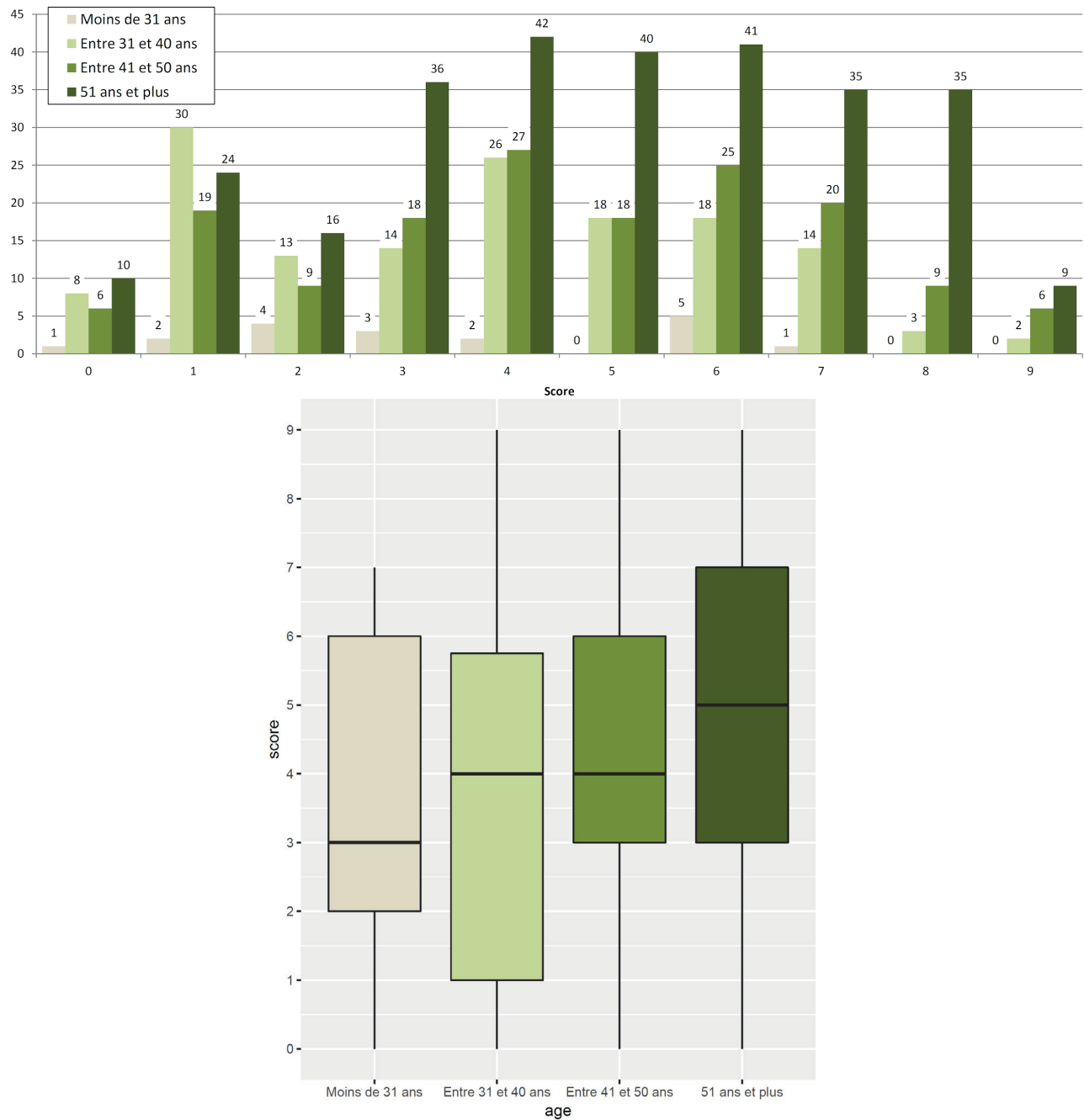


Tableau 51 : Âge des répondants

	Moins de 31 ans	Entre 31 et 40 ans	Entre 41 et 50 ans	51 ans et plus
Moyenne	3,56	3,7	4,49	4,81
1 ^{er} quartile	2	1	3	3
Médiane	3	4	4	5
3 ^{ème} quartile	6	5,75	6	7
Écart-type	2,15	2,3	2,37	2,36

ANOVA
p=2.69E-05

Parmi nos répondants, les plus âgés ont de meilleurs scores.

Figure 4 - Score selon ancienneté d'installation

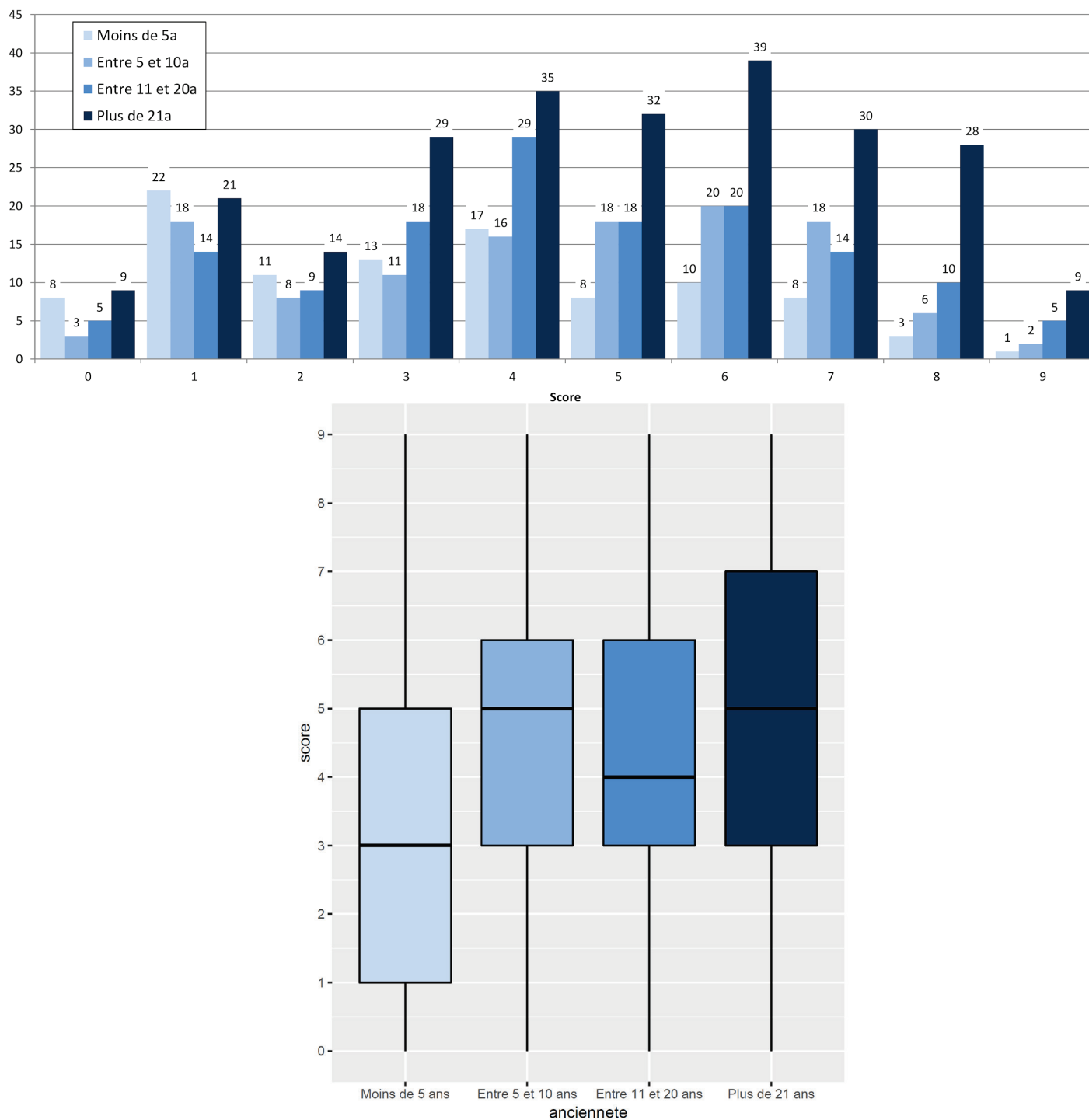


Tableau 52 : Score selon ancienneté d'installation

	Moins de 5 ans	Entre 5 et 10 ans	Entre 11 et 20 ans	Plus de 21 ans
Moyenne	3,37	4,44	4,47	4,82
1 ^{er} quartile	1	3	3	3
Médiane	3	5	4	5
3 ^{ème} quartile	5	6	6	7
Écart-type	2,31	2,31	2,29	2,38

ANOVA
p=5.33E-06

Parmi nos répondants, ceux dont l'installation figure parmi les plus anciennes avaient de meilleurs scores de sauvegarde.

Figure 5 - Score selon type d'exercice

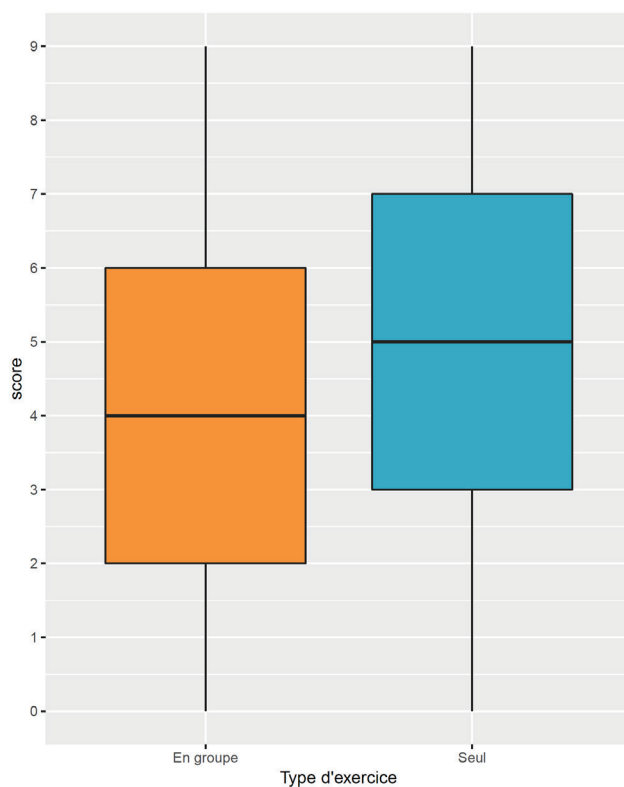
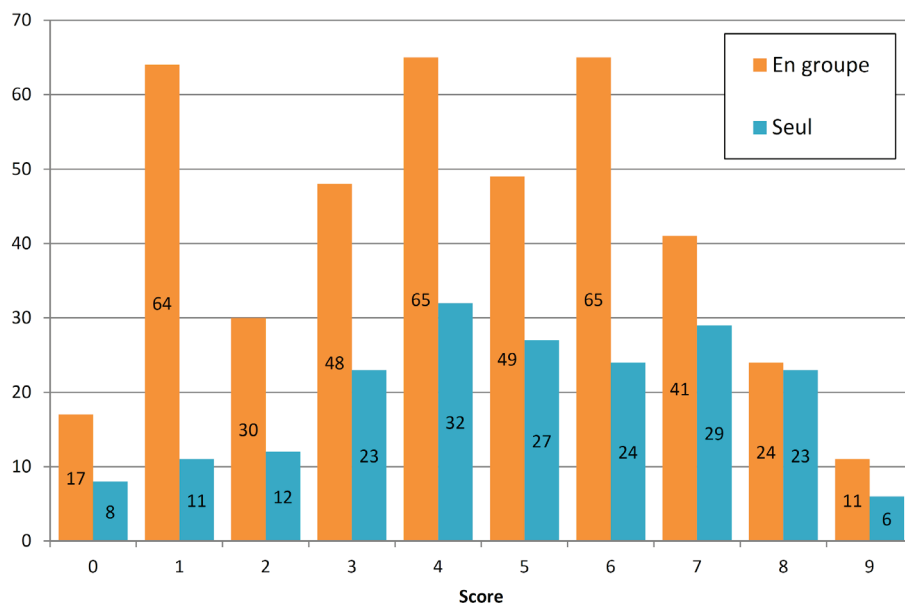


Tableau 53 : Score selon type d'exercice

	En groupe	Seul
Moyenne	4,21	4,88
1 ^{er} quartile	2	3
Médiane	4	5
3 ^{ème} quartile	6	7
Écart-type	2,38	2,32

Test t de Student
p=0.000961

Parmi nos répondants, ceux exerçant seuls avaient de meilleurs scores de sauvegarde.

Les facteurs d'étude qui ont une influence statistiquement significative ($p < 0,05$) sur le score de sauvegarde ont été le sexe, l'âge, l'ancienneté d'installation et le type d'exercice des répondants. L'analyse des autres caractéristiques démographiques n'a pas retrouvé de lien statistiquement significatif.

Pour éliminer les facteurs de confusion entre ces quatre critères démographiques, significatifs en univarié, nous avons réalisé une analyse multivariée par ANOVA (résultats complets reportés [en annexe](#)) :

Tableau 54 : Résultats ANOVA

âge	p=1.49E-05
sexe	p=4.14E-06
groupe	p=0.05639
ancienneté	p=0.03356

Au final, l'analyse ANOVA multivariée confirme qu'un **lien statistiquement significatif** existe bien **entre le score de sauvegarde et l'âge, le sexe et l'ancienneté d'installation**, mais pas le type d'exercice.

4. Discussion

4.1. Démographie des répondants

Nous pouvons comparer nos résultats aux statistiques fournies par le CNOM sur la démographie médicale de 2017 concernant spécifiquement les généralistes [\(40\)](#).

- Sex-ratio

Le sex-ratio de la population des généralistes français en activité de 2018 était de 42328 femmes pour 45473 hommes, soit 0,93. Celui de nos répondants était de 276/333 soit 0,829. La majorité d'hommes est proche dans les deux populations.

Il est **possible** qu'il y ait ici un **biais de sélection**. Les hommes sont possiblement plus attirés par l'usage de la technologie informatique que les femmes, d'où un sex-ratio plus faible que celui de l'ensemble des généralistes.

- Répartition en classe d'âge

La pyramide des âges obtenue n'est pas strictement comparable à celle fournie par l'ordre des médecins, car la subdivision en classes d'âge de notre questionnaire est différente de la leur. Notre question demandait aux répondants de se positionner dans une classe d'âge (ne pas noter son âge à l'année près). Ce souci de bienséance ne nous permet donc pas de repositionner nos classes d'âge pour correspondre à celles de l'Atlas de démographie médicale du CNOM.

On remarque cependant que la répartition globale entre notre population de répondants et celle des généralistes français est grossièrement comparable, avec une majorité d'hommes de plus de 50 ans.

Figure 6 - Âge des répondants

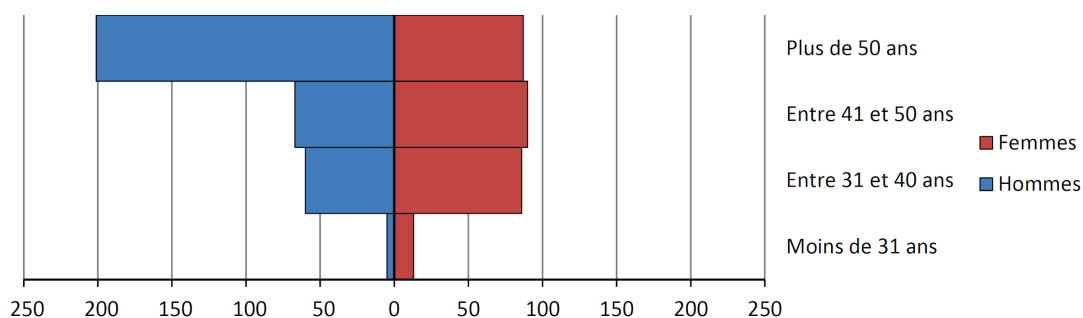
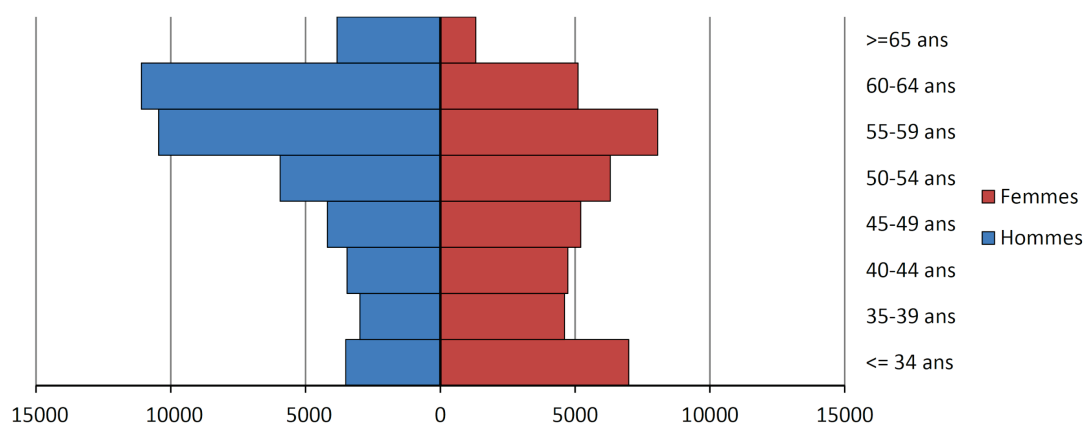


Figure 7 - Âges des généralistes français en activité en 2018 (40)



Les résultats plus détaillés du CNOM permettent d'extraire la médiane d'âge des généralistes français, qui se situe entre 50 et 54 ans. La médiane d'âge de nos répondants était plus basse, entre 41 et 50 ans. Notre population n'est donc pas strictement superposable à celle de tous les généralistes.

- Ancienneté d'installation

La majorité des répondants est installée depuis au moins plus de 10 ans, avec une médiane située entre 11 et 20 ans. Encore une fois, un questionnement en classe d'âge et non en chiffre entré à la main empêche d'être plus précis.

Nos subdivisions en classe d'âge et de durée d'installation sont différentes : elles sont espacées de 10 ans pour les classes d'âge et de 5 ans pour la durée d'installation. On gagne donc en granularité sur l'ancienneté d'installation par rapport à l'âge. Ces deux variables peuvent être indépendantes, mais ont de fortes chances d'être intriquées. Un généraliste peut être âgé et s'être installé récemment (changement de lieu d'exercice), mais tous ne seront pas dans ce cas et un âge plus élevé ira de pair avec une installation plus ancienne.

- Territoire et mode d'exercice

Les participants exerçant en groupe sont 2,12 fois plus nombreux que ceux en solo. La proportion de médecins exerçant en groupe tendrait (41) à être plus importante au fil des ans. Nous ne disposons pas de chiffres précis pour les comparer à ceux de la démographie des généralistes en France.

Ainsi, un nouveau **biais de sélection** potentiel émerge car notre population de répondants serait différente de la population générale. La majorité des répondants étant des hommes dans les classes d'âge supérieures, nous devrions nous attendre à davantage de praticiens exerçant seuls.

Le mode d'exercice le plus fréquent était l'exercice en groupe (67,98%), et le territoire d'exercice le plus représenté est urbain (63,55%). L'exercice en ville va de pair avec un accès à Internet plus performant. L'hypothèse serait que les répondants urbains soient de ce fait plus sensibles à la problématique informatique.

Les bénéfices des nouvelles technologies et méthodes de sauvegarde sont possiblement sous-représentés chez nos répondants. Premièrement, les jeunes médecins, à l'exposition potentielle plus précoce aux technologies informatiques, sont minoritaires parmi les répondants.

Ensuite, 20% des répondants sont installés depuis moins de 5 ans. Une installation récente a de grandes chances d'utiliser les dernières versions des progiciels, et donc d'utiliser les dernières avancées informatiques proposées aux généralistes. Nos résultats devraient donc laisser transparaître des résultats en lien avec l'adoption de méthodes informatiques récentes pour cette frange de population.

- Pratiques d'usage informatique et du progiciel

Plus de 60% de nos répondants fonctionnent en architecture client/serveur. Parmi les 414 généralistes exerçant en groupe, la grande majorité (371 soit 89,6%) partage ses données avec d'autres membres du groupement. De fait, ces répondants échangent de l'information médicale sur un réseau local. Il s'agit d'un premier pas sur le chemin d'un export de données hors des murs du cabinet médical.

Ne pas centraliser les données de santé au sein du progiciel oblige à sauvegarder les données non incluses par un autre moyen. Ceci ajoute des étapes et des problèmes potentiels au processus de sauvegarde. La proportion de nos répondants ne stockant pas l'intégralité de leurs données dans leur progiciel est plus faible dans nos répondants (106 soit 17,41%) que les 36% du Dr STAECHELÉ en 2012. Parmi les 117 répondants réalisant une sauvegarde et ne stockant pas l'intégralité de leurs données de travail dans le progiciel, seule la moitié (51,28%) rapporte inclure ces données dans leur sauvegarde. Le nombre de répondants dans la situation où une partie des données de travail ne sont pas sauvegardées représente une minorité des répondants : 16 soit 2,63% des 609 répondants. Ces pratiques ne sont pas acceptables car des risques importants de perte de données sont pris alors que les méthodes de protection sont présentes mais non utilisées.

- Existence d'une sauvegarde

Dans les précédents travaux de thèse, le taux de non réalisation de sauvegarde était de 1% en 2007 pour le Dr BRAHMI, puis en 2012 moins de 5% pour le Dr CHERIET et 13% pour le Dr STAECHELÉ. La grande majorité de nos répondants (527 soit 86,54%) déclare avoir au moins une sauvegarde. La question spécifiait ne pas tenir compte des éventuelles sauvegarde des progiciels en *cloud*. En effet, les éditeurs de progiciel en *cloud* sont théoriquement obligés de réaliser des sauvegardes de leur côté, mais le médecin répondant ne peut le certifier pour eux. Nous considérons dans la suite de ce travail que les utilisateurs de progiciel *cloud* ont au moins une sauvegarde gérée de cette façon.

Le nombre réel de participants ne réalisant absolument aucune sauvegarde émerge plus loin dans les résultats, après recoupement de l'utilisation de progiciel en *cloud* et du nombre d'exemplaires de sauvegarde : 12 répondants, soit 1,97% du total ([chiffre en rouge dans le tableau 40](#)), ce qui correspond à un **taux de sauvegardes réalisées à 98,03%**. Notre taux d'absence de sauvegarde est

bien plus faible que ceux des travaux de 2012, mais a interrogé un échantillon de population bien plus large.

Plus préoccupante est notre proportion de médecins ne sachant pas s'ils ont une sauvegarde de leurs données ou non (21 soit 3,45%), même si elle reste faible. Si l'on considérait que ne pas savoir si on dispose d'une sauvegarde équivaut à ne pas en avoir, le nombre d'absence de sauvegarde grimperait à 33 soit un taux de 5,41%. Dans tous les cas, le taux d'absence de sauvegarde de nos répondants n'est pas plus faible que le meilleur des résultats des thèses précédentes.

4.2. Ressenti et connaissances des répondants

- Responsabilité ressentie du médecin généraliste envers ses données

La question posée était exactement *“Selon vous, le médecin généraliste est-il responsable de la sécurité de ses données informatisées s’il ne fait appel à aucun prestataire ? s’il en confie la gestion à un prestataire ?”*.

Selon l'Ordre des Médecins, en cas d'altération ou de perte de données, la personne responsable serait *“le médecin ou l'établissement de santé s'il assure lui-même la conservation des données ou l'hébergeur agréé si les données de santé lui ont été confiées.”* [\(35\)](#). Les répondants semblent bien conscients de la responsabilité du médecin hébergeant lorsqu'il est seul impliqué, à une écrasante majorité (95%).

Le report de responsabilité en cas de perte à l'hébergeur contractuel semble moins clairement compris : 68,3% pensent que le généraliste reste responsable de ses données si confiées à un prestataire, avec 18% de répondants incertains à ce sujet.

- Connaissance de la localisation des données de travail

Parmi les 503 participants n'ayant pas de progiciel en *cloud*, seuls 44% sont capables de localiser leurs données de travail sur le disque dur de la machine hébergeante. Ce résultat reste cohérent avec les 48% du Dr STAECHÉLÉ. Cette méconnaissance d'emplacement virtuel est un obstacle potentiel à la réalisation de sauvegardes manuelles des données, qui nécessite de sélectionner l'emplacement des données sources à sauvegarder.

4.3. Pertes de données

On ne s'est concentré ici que sur les 5 dernières années pour plusieurs raisons. D'abord, éviter d'être redondant avec la période d'avant 2012, date qui correspond aux données des deux thèses précédentes les plus récentes. Ensuite, se limiter à une période récente limite le biais de mémoire.

- Existence d'une perte de données au cours des 5 dernières années

La proportion de répondants ayant subi une perte (111 soit 18,23%) peut sembler faible, mais étant donné la courte période considérée, il est probable qu'il y ait eu d'autres pertes antérieures chez les

généralistes les plus âgés. Les répondants sont majoritairement en fin d'exercice, cela pourrait suggérer qu'un exercice déjà long n'est pas forcément protecteur des pertes de données.

Comme l'avait remarqué le Dr BRAMI alors en 2007, il existe encore une discordance entre la confiance des médecins en la sécurité de leurs données et les taux de perte constatés. A l'époque, 31% des répondants avaient déjà subi une perte de données. En 2012, les taux de pertes déclarées étaient de 45% pour le Dr STAECHÉLÉ, mais la question était posée sans borne temporelle. On ne peut donc pas conclure à une baisse du taux de perte, mais seulement constater la persistance de pertes de données.

- Causes des pertes de données

Parmi nos 111 répondants victimes de perte de données, une cause est identifiée pour 103 (92,79%) des cas. Le dysfonctionnement logiciel seul est désigné comme responsable n°1 à 30,63%, puis la panne matérielle seule à 28%, et enfin l'erreur humaine seule en 3ème position à 12,61%. 15 (13,51%) des répondants déclarent deux causes simultanées à leur perte de données, et parmi eux 5 (33,33%) mentionnent l'erreur humaine. Les autres causes rapportées sont minoritaires. Comme expliqué dans l'introduction, l'erreur humaine serait dans de nombreuses entreprises sous-estimée comme cause potentielle de perte de données, car retrouvée comme cause principale dans 20 à 30% des pertes de données. Il est possible que nos répondants soient dans cette situation.

On remarque une absence totale de vol de support de données déclaré responsable de perte de données. Dans les travaux précédents, notamment en 2012 dans le travail du Dr STAECHÉLÉ, était rapportée une hausse des taux d'effractions et de vol de matériel informatique. Cependant, notre question était tournée autrement : elle ne demande pas s'il y a eu effraction ou vol, mais si le vol était responsable de perte de données. Ainsi, il nous est impossible de savoir si les répondants ont été sujets à des effractions, ce qui sort de notre travail de thèse. On peut par contre supposer que même s'il y a eu pour certains vol d'un support de données, des sauvegardes multiples leur ont permis au final de ne pas avoir de perte de données, et donc de ne pas mentionner de vol dans cette question.

- Quantité de données perdues

On remarque que la perte de données initiale déclarée est majoritairement courte : 82,9% ont perdu moins d'un mois de travail en premier lieu. Parmi les répondants ayant perdu des données, la très large majorité (95,5%) était protégée par un système de sauvegarde. Ceci explique très probablement le fait qu'aucune perte définitive de données de plus de 6 mois n'a été déclarée.

Cependant, **seule la moitié (51,9%) des participants affectés par une perte de donnée ont pu restaurer au moins une partie de leurs données**, l'autre moitié n'ayant pas pu utiliser leur système de sauvegarde, ce qui montre bien qu'un système de sauvegarde en place peut ne pas être efficace. De plus, **seul 27,4% de ceux ayant perdu des données ont pu en récupérer la totalité**. Au final, **au cours des 5 dernières années, sur le total des répondants, 51 sur 609 soit 9,03% ont perdu des données définitivement, et ce alors que la très grande majorité des affectés était alors**

protégée par un système de sauvegarde (plus de 95%). L'efficacité de ces systèmes a été suboptimale pour plus de la moitié des affectés.

Par rapport au travail de 2012 du Dr STAECHELÉ, les quantités de données perdues définitivement coïncident : une moyenne de 6,35 jours et 80% de perte inférieure à une semaine par rapport à 72% en 2012. Notre taux de perte définitive de 9% est aussi proche des 14% de 2012.

Une perte finale moyenne de moins d'une semaine de travail peut sembler être un résultat satisfaisant, mais les capacités de l'informatique devraient théoriquement permettre une restauration totale sans erreur. Une perte de données médicales reste à notre avis difficilement acceptable au vu des outils existants.

Dans les réponses sur la quantité de données définitivement perdues, trois réponses ont été de "0 jour", qui est une réponse illogique puisque cette question n'était posée qu'à ceux ayant pu restaurer toutes leurs données mais pas en totalité. La réponse "0 jour" aurait dû empêcher la validation de la réponse. En l'état, ces réponses nous montrent donc une possible incompréhension de la notion de "perte définitive" pour ces trois répondants.

- **Moyen de restauration des données après la perte**

Les réponses obtenues à la question "*quelles sauvegardes ont été utilisées pour restaurer les données perdues ?*" ont été surprenantes, dans le sens où tous les concernés disent n'avoir eu recours à une sauvegarde ni locale ni externe. Ces résultats incohérents nous ont fait revenir sur le fichier de structure logique du questionnaire. La condition d'apparition de la question "par quel moyen a été faite la restauration" était d'avoir répondu "oui" à la question "avez-vous pu restaurer les données à partir d'une sauvegarde".

Après vérification, elle a été programmée correctement, mais pour une raison que l'on ne s'explique pas elle n'apparaît pas dans les questions posées, même si les conditions d'affichage sont remplies. Nous avons donc été victimes ici d'un **dysfonctionnement logiciel** sur le fichier de logique du questionnaire, possiblement spécifique à notre version de Limesurvey.

- **Temps pour retrouver un fonctionnement normal après la perte**

La médiane à 2 jours diffère beaucoup de la moyenne à 28,47j : 77% des généralistes ayant perdu des données sont retombés sur leurs pieds en moins d'une semaine, le reste a rapporté des durées très éparpillées. Il nous semble difficile d'accepter des durées de perturbation supérieures à quelques jours lors d'une perte de données avec restauration bien conduite.

4.4. Méthodes de sauvegarde

- **Automatisation des sauvegardes**

L'**automatisation** a pour objectifs d'apporter un gain de temps à long terme et une régularité dans la réalisation de tâches répétitives. Elle nécessite une mise en place adéquate, et une surveillance des

résultats obtenus avec une première vérification dès la mise en place. Par la suite, une surveillance ponctuelle de son bon fonctionnement est recommandée.

Dans nos résultats, 291 soit 55,2% des sauvegardants ont automatisé leur processus de sauvegarde. Les 43% qui n'ont pas automatisé leur sauvegarde réalisent donc leurs sauvegardes manuellement. Une sauvegarde manuelle peut donner une impression de contrôle à l'utilisateur qui a un retour visuel au moment de l'exécution, contrairement à une sauvegarde automatique en arrière-plan. L'utilisateur perd du temps pour aller lancer l'opération à chaque fois, et risque même d'oublier de la lancer. Ce mode de fonctionnement est non souhaitable car source d'erreurs et d'irrégularité : un intervalle de temps non couvert par une sauvegarde augmente la quantité de données perdues potentielle.

Une **sauvegarde automatisée quotidienne** nous semble être un **objectif minimum**. La perte d'une journée de saisie est déjà handicapante et difficilement acceptable quand les technologies actuelles permettent une sauvegarde incrémentale rapide. De plus, automatiser les sauvegardes équivaut sur de nombreux systèmes modernes à la création d'[instantanés](#) entre les sauvegardes proprement dites. Il devient alors possible de restaurer un système avec une précision inférieure à une journée. De faible taille, les instantanés sont créés rapidement, avec un impact en performance sur le poste de travail minimal.

En plus d'être quotidienne, l'automatisation de la sauvegarde doit si possible être programmée hors des heures ouvrées. Il n'est pas conseillé d'automatiser sa réalisation sur les heures ouvrées, car cela pourrait perturber le fonctionnement du progiciel, voire empêcher le généraliste de travailler si le processus de sauvegarde interdit de nouvelles entrées pendant son travail.

- Nombre d'intervenants dans la réalisation ou automatisation de la sauvegarde

81,6% des sauvegardants sont dans une situation de gestion à une personne. Elle permet de limiter le risque d'erreur humaine, mais limite la résilience du système en cas de l'absence de son gestionnaire. Les généralistes sauvegardants, seuls ou en groupe, gérant leur système seuls représentent 43,5% des cas.

L'implication du généraliste dans la gestion de sa sauvegarde est primordiale. Cependant, 39,85% de nos sauvegardants ne sont pas impliqués personnellement dans leur processus de sauvegarde. Ceci signifie qu'**en cas de panne** ou de besoin de restauration, s'ils ne connaissent pas les procédures à suivre, **ces généralistes peuvent être totalement dépendants d'un tiers pour poursuivre ou reprendre leur activité**. On peut l'affirmer pour la majorité d'entre eux : 176 (83,81%) disent ne pas connaître le plan de continuité, et 179 (85,24%) disent ne pas connaître le plan de restauration.

- Utilisation de l'outil de sauvegarde intégré au progiciel

L'outil intégré au progiciel est la méthode la plus fiable pour sauvegarder les données qu'il contient : le progiciel et son outil de sauvegarde sont créés, mis à jour et maintenus par le même éditeur. Celui-ci est censé se charger d'assurer une synergie parfaite entre eux, et ce à chaque mise à jour. La résilience du système de sauvegarde en est donc augmentée.

Il est tout à fait possible de sauvegarder efficacement par une méthode autre que celle prévue dans le progiciel, mais la mise en place en est moins aisée et surtout moins sûre qu'un processus intégré. Lors de l'utilisation d'un programme tiers de sauvegarde, la synergie progiciel et méthode de sauvegarde n'est pas assurée : rien ne garantit que l'enchaînement d'opérations paramétrées restera valide lors de chaque mise à jour du progiciel. Si la structure ou l'emplacement des données à sauvegarder sont modifiées, un programme tiers pourrait continuer à fonctionner sans s'adapter aux changements du progiciel. On obtiendrait alors des sauvegardes incorrectes, donc potentiellement inexploitable et inutiles. Parmi les sauvegardants, 129 (21,18%) sont dans cette situation. Les méthodes tierces sont variées : l'emploi d'un programme tiers est en tête, suivi de la sauvegarde manuelle et d'un outil intégré au système d'exploitation. Plus étonnant, 15 utilisent deux voire trois méthodes tierces surajoutées, ce qui multiplie les risques d'erreur.

De ces trois méthodes, la copie manuelle est la plus à risque : sans une connaissance précise de la structure des données sauvegardées et du lieu de stockage des paramètres progiciels, il est très facile d'aboutir à une sauvegarde non restaurable. Si le progiciel utilise un système de base de données pour organiser les données, il faut à la fois dupliquer les données de la base et les paramètres du système de gestion de la base de données. Ainsi, se limiter à une copie du contenu de la base de données ne suffira pas à assurer un bon fonctionnement après restauration. L'utilisation de programmes tiers ou intégrés au système d'exploitation peut être un peu moins à risque, mais doit également être bien paramétrée pour fonctionner.

Dans tous les cas, un test de restauration reste indispensable pour valider le bon fonctionnement de chaque méthode particulière. Or, **pour 89 des répondants, leur méthode de sauvegarde non progicielle n'a jamais fait l'objet d'un test de restauration.**

- Chiffrement des sauvegardes sur support physique

Le chiffrement des données est accessible à tous les utilisateurs d'ordinateurs modernes depuis plus de 10 ans. Il est protecteur dans les situations où les données sont séparées de leur clé de chiffrement, c'est-à-dire en cas de vol de support physique ou d'export chez un tiers non agréé HDS. Cette **problématique est mal connue des répondants** : malgré l'explication claire du principe du chiffrement, **plus de la moitié des sauvegardants ne savent pas si leurs sauvegardes sont chiffrées. Plus d'un quart savent ne pas chiffrer leurs sauvegardes.** Les résultats de nos répondants sont cependant meilleurs que ceux de 2012 du Dr STAECHÉLÉ : à l'époque, aucun des généralistes interrogés n'employait de chiffrement.

Ce procédé est relativement technique dans son principe, mais il est simple à mettre en œuvre sur le poste de travail hébergeant les données. Il devrait donc faire partie d'un processus de sauvegarde bien conduit. Le chiffrement peut intervenir à plusieurs niveaux : par le système d'exploitation, par le progiciel lors de la sauvegarde, par le support physique cible. Dans tous les cas, il est à réaliser avant tout export de données.

- Utilisation d'un *cloud* tiers comme support de sauvegarde

Nous avons cité dans la règle des 3-2-1 la nécessité théorique de conserver une copie de sauvegarde hors-site, ce à quoi répond un *cloud* tiers. Parmi les sauvegardants, seuls 10,6% emploient un tel système. Parmi eux seuls 7 l'emploient en complément de leur progiciel déjà en *cloud*, ce qui est une excellente pratique (en 3-2-2), bien que coûteuse.

Etant donné qu'un *cloud* tiers n'est pas forcément agréé HDS, le chiffrement avant envoi permet de réduire le risque de lecture et d'exploitation des données par l'hébergeant. Pour les sauvegardants sur *cloud* tiers, seuls 15 sont certains de chiffrer les données avant envoi au *cloud* tiers. 25 autres ne savent pas si leurs données sont chiffrées lors de cet export. Dans ces situations, il y a incertitude quant à la confidentialité des données confiées au service tiers.

On pourrait concéder cette méconnaissance du sujet s'il n'y avait aucune situation d'externalisation de données non chiffrée vers un tiers. Malheureusement, c'est le cas de façon certaine pour 5 de nos répondants. On ne sait cependant pas si ces 5 répondants envoient leurs données sur un *cloud* agréé HDS.

Un sauvegardant dit envoyer une sauvegarde sur un *cloud* tiers sciemment non chiffrée avant envoi, ce qui pourrait être inquiétant.

- Nombre d'exemplaires et types de supports de sauvegarde

Le principal objectif de cette question est de cerner si la [règle des 3-2-1](#) est respectée pour chacun des répondants. Les résultats ne tiennent pas compte de l'exemplaire original des données que sont les données de travail du généraliste, puisque l'on demandait le "nombre de sauvegardes par support physique". Les répondants ne devraient pas avoir comptabilisé leurs données de travail comme un exemplaire de sauvegarde dans leurs réponses. En effet, la définition d'une sauvegarde était détaillée au début du questionnaire comme étant "*[une] copie intégrale des données de travail, réalisées à un instant T, stockées sur un support différent de celui des données de travail.*"

Puisque la médiane d'exemplaires totaux de sauvegarde est à 2 sur l'ensemble des répondants, on peut affirmer qu'**au moins la moitié des répondants disposent de 3 exemplaires de leurs données de travail et donc respectent le critère "3" de la règle des 3-2-1.**

Sur l'ensemble des répondants, la moyenne du nombre de supports totaux est à 1,88, donc encourageante car proche du critère "2" de la règle des 3-2-1. Si l'on ne tient compte que des répondants en installation locale pure (aucun *cloud* progiciel ou tiers), la moyenne se rapproche encore de 2.

Sur l'ensemble des répondants, **376 (61,74%) disposent d'au moins un exemplaire de sauvegarde sur support externalisé, et donc respectent le critère "1" de la règle des 3-2-1.** C'est mieux que les 47% rapportés par les répondants du Dr STAECHELÉ en 2012.

Concernant le cas particulier des utilisateurs de progiciel *cloud* et leur relatif manque de supports physiques associés : ceci peut être expliqué par la difficulté d'automatiser un export des données du

progiciel *cloud* en dehors de ses serveurs. Cette opération devrait être possible en routine, mais est encore bien souvent limitée par l'éditeur du progiciel. Elle peut être réalisable sur demande du généraliste si son progiciel est labellisé [“e-santé Logiciel Maisons et Centres de santé”](#).

De plus, le généraliste employant un progiciel en *cloud* peut ne pas ressentir le besoin de réaliser des sauvegardes sur un support physique, si la sécurité apportée par les méthodes du prestataire est perçue comme suffisante. Selon le niveau de sécurité garantie par le fournisseur de service, il peut être prudent d'éviter que tous les exemplaires de sauvegarde ne soient centralisés par un seul opérateur. Seule une situation où le fournisseur de service est agréé HDS autoriserait l'absence d'exemplaires de sauvegardes sur supports physiques gérés par le généraliste.

Les méthodes de sauvegarde les plus fréquemment utilisées en installation locale sont perfectibles, car ne respectant pas les trois éléments de la [règle des 3-2-1](#). On notera également un emploi large de disque dur externalisé. Comme expliqué dans la définition de [pérennité des supports](#), un disque dur classique a l'avantage du coût et de la fiabilité en usage statique, mais n'est pas le support idéal pour un transport physique de données car contenant de nombreuses pièces mécaniques de précision. Un transport itératif, avec possibles chocs à chaque voyage, peut réduire la durée de vie du support et le rendre inutilisable. Les supports d'export de choix sont les systèmes à base de mémoire Flash, donc SSD portables ou clés USB.

Les clés USB peuvent être un support massif mais avec taux de transfert relativement lent, d'autant plus qu'on transfère de petits fichiers. Cette lenteur n'est pas insurmontable si l'on réalise cette opération en arrière-plan. Les SSD sont plus récents, permettent une sauvegarde rapide, et un transport hors du cabinet avec risque de corruption de données plus faibles qu'un disque dur standard, mais à un coût plus élevé.

- Vérification du processus de sauvegarde

Plus de la moitié des répondants (57%) répondent vérifier la réussite de leur processus de sauvegarde, et parmi eux près de 75% peuvent justifier leur méthode de vérification.

Une **pratique idéale** serait une **double vérification** : **existence des fichiers** de sauvegarde sur le support de destination (par un explorateur de fichier), et **contenu du journal de sauvegarde** créé par le logiciel de sauvegarde. Une vérification du journal de sauvegarde seule peut être considérée suffisante, mais tous les progiciels ne disposent pas forcément de la fonction de journalisation. Si elle existe, elle est à activer. Si elle n'existe pas et que l'utilisateur en ressent le besoin, ce peut être une des raisons poussant à changer de logiciel de sauvegarde.

Se contenter de vérifier si des fichiers ont été créés ou modifiés n'est pas suffisant pour être certain de la réussite de la sauvegarde, car si une erreur survient au sein du logiciel de sauvegarde, elle sera notifiée dans son journal mais pas forcément visible en consultant la liste des fichiers. Il est possible que le logiciel avertisse l'utilisateur d'une erreur dans une fenêtre bien visible au premier plan sur l'écran, mais cela dépend du logiciel employé et n'est pas obligatoirement le cas par défaut.

Parmi les sauvegardants, 43% disent ne pas vérifier la réussite du processus de sauvegarde. Cette pratique est améliorable. Le progiciel doit être équipé de cette fonction, qui devra être activée. Elle n'a pas à être systématique, et devrait être faite au moins sur les premières sauvegardes puis de temps en temps. S'assurer que le logiciel sauvegardant est paramétré pour signaler une erreur automatiquement à l'utilisateur permet de gagner du temps sur cette manœuvre.

4.5. Restauration de sauvegarde

Pour vérifier le bon fonctionnement d'un système de sauvegarde, il est nécessaire de réaliser des **tests de restauration** : repartir d'un progiciel avec une base de données vide et utiliser la sauvegarde la plus récente pour retrouver toutes ses données de façon exploitable et fonctionnelle au sein du progiciel.

Le premier test de restauration permet de vérifier la réussite du paramétrage des logiciels impliqués, et dans le cas d'un logiciel tiers de sauvegarde si la méthode employée est la bonne. Ce **premier test** est **absolument indispensable** : vérifier uniquement la création de fichiers de sauvegarde ne permet pas de savoir si le progiciel pourra effectivement utiliser ces données une fois restaurées.

Près de 75% des répondants n'ont jamais réalisé de test de restauration. Ce chiffre n'a pas évolué depuis 2012 par rapport aux 72% relevés dans le travail du Dr STAECHÉLÉ.

Cette vérification peut prendre un peu de temps, et doit être détaillée dans le mode d'emploi du progiciel donc facile à effectuer. Étant donné qu'elle se fait lors du paramétrage de la sauvegarde, l'opérateur est déjà engagé sur une opération chronophage : il a tout intérêt à valider alors la bonne marche de sa sauvegarde par ce test. Une nouvelle installation progicielle aura une base de données vide. Il est pertinent de créer un dossier patient test ou d'importer quelques anciens dossiers à ce moment-là. La base de données sera alors de petite taille. Les opérations de sauvegarde puis de restauration seront testables rapidement.

Les tests ultérieurs permettent de vérifier que la méthode mise en place reste efficace, malgré l'évolution éventuelle du progiciel. **Un test de restauration itératif est le seul procédé qui permet de valider le bon fonctionnement d'un système de sauvegarde et de garantir son utilité en cas de besoin.**

Concernant notre travail de thèse avec la sauvegarde de la base de données du questionnaire, il n'y a jamais eu d'incident ni de besoin de restauration de données. Nous avons utilisé [Duplicati](#), un utilitaire qui fonctionne bien pour des sauvegardes de petites tailles (inférieure au Go, moins d'une centaine de fichiers).

Pour illustrer notre propos, un exemple personnel : ce même outil, Duplicati, a été utilisé pour la sauvegarde des données personnelles du thésard sur une autre machine. La quantité de données sauvegardée était bien plus importante (plusieurs centaines de Go, centaine de milliers de fichiers). Un premier test de restauration s'est bien déroulé. Après plusieurs mois de sauvegarde, un nouveau

test de restauration a déclenché une alerte par Duplicati avec plusieurs erreurs signalées. Plusieurs fichiers restaurés étaient corrompus une fois restaurés, et l'un d'entre eux était illisible. Ce type de dysfonctionnement n'est pas acceptable dans un système de sauvegarde informatisé, théoriquement capable d'une recopie sans erreur. Dans ce cas, l'incident a conduit à revoir la méthode de sauvegarde. Le logiciel Duplicati n'étant pas adapté, il a été changé au profit d'un autre (Borg-Backup, également gratuit mais notoirement plus fiable).

Il est nécessaire pour les éditeurs de logiciel de proposer une solution éprouvée et fiable de sauvegarde. Dans le cas où un généraliste aventureux choisit de réaliser ses sauvegardes par ses propres moyens, il devra absolument réaliser des tests de restauration approfondis.

Dans le cadre de l'exercice professionnel des généralistes, le Portail d'Accompagnement Cybersécurité des Structures de Santé spécifie que "*à minima, un **test annuel [de restauration]** doit être réalisé.*" [\(18\)](#), ce que **seuls 53 répondants (soit 10,06% du total)** disent faire. Un investissement annuel de temps dédié à cette opération est donc conseillé.

Près des **deux tiers des participants** ayant réalisé un **test de restauration** l'ont réalisé **uniquement sur une machine séparée de la base de production**. Cette méthode est la bonne car ne risquant pas de provoquer une corruption de la base de production en cas d'accident ou de mauvaise manipulation.

En pratique, une sauvegarde doit être restaurée sur une machine où un exemplaire du progiciel sauvegardé est présent. Bien souvent, les progiciels sont vendus à la licence, avec une licence par poste, voire par utilisateur. Il est conseillé de restaurer sur une machine hors base de production, c'est-à-dire dédiée aux opérations informatiques de maintenance : il en résulte un surcoût potentiel d'une licence. Etant donné l'importance de cette opération, **les éditeurs de progiciel devraient fournir une licence supplémentaire gratuite dédiée à cet usage**. Parallèlement, le généraliste restaurant ne devra pas faire employer cette licence supplémentaire pour d'autres usages, comme le travail quotidien d'un praticien supplémentaire.

4.6. Conduite en cas de panne

En cas de dysfonctionnement ou perte de données, afin de permettre la reprise d'activité ou restauration de données rapidement, il est intéressant de mettre en place un **plan de continuité** et un **plan de restauration**. Ces plans devraient être établis lors de la mise en place du système informatique, avec les généralistes qui doivent les connaître pour pouvoir les appliquer.

Sur les 609 répondants, 179 (29,39%) connaissaient le plan de continuité à suivre en cas de panne, et 194 (31,86%) connaissaient le plan de restauration. Ces chiffres contrastent avec les bons résultats obtenus dans la section sur les pertes de données. Il est possible que les médecins ayant perdu des données aient fait appel à un tiers qui connaissait ces plans de continuité et de restauration. Ceci suggère que les **répondants** ne sont **pas autonomes** dans la gestion des pannes, malgré le faible nombre de pertes de données signalé.

4.7. Perception du risque informatique

L'évaluation complète du degré de sécurité informatique ne faisait pas partie de notre travail de thèse, mais la qualité d'une sauvegarde fait partie intégrante de leur sécurité. Au vu des autres réponses et des scores de sauvegarde, il existe clairement encore une **discordance entre confiance accordée à la sécurité des données et la réalité des méthodes employées** : une moyenne de 3,6/5 en confiance est en décalage par rapport à la moyenne des scores de sauvegarde à 4,42/10.

La confiance de nos répondants en leurs compétences personnelles est d'autant plus élevée que l'on évoque une tâche réalisée fréquemment : on côtoie son poste et ses données de travail au quotidien, alors que la sauvegarde n'est pas forcément quotidienne. Le test de restauration l'est encore moins, puisque jamais fait pour presque 75% des répondants.

En revanche, nos répondants semblent conscients des faiblesses de leurs méthodes de restauration. Nous avons vu que les tests de restauration sont rares parmi nos répondants, et que parmi ceux ayant perdu des données seule une fraction a réussi à les restaurer au moins partiellement à partir d'une sauvegarde.

Sur le terrain, les systèmes de sauvegardes peuvent être gérés selon une politique de "aussitôt réglé, aussitôt oublié". Une absence de vérification du bon fonctionnement aboutit à une méconnaissance de l'état du système, mais aussi à l'absence de pratique des méthodes de premier recours en cas de perte de données.

Les généralistes sont formés pendant leurs études aux gestes de premiers secours, avec des formations de rappel proposées au cours de congrès et formations. En revanche, proposer à un généraliste de refaire le point sur les procédures de restauration de sauvegarde ne fait pas partie des pratiques à rafraîchir. Dans le cas d'une gestion autonome de ses sauvegardes, c'est un point absolument crucial. Si la gestion des sauvegardes est partagée, il faut s'assurer qu'au moins une personne en charge du système se tienne à jour des méthodes à employer.

4.8. Archives

Cette section ne concerne que les 157 participants réalisant une archive informatique de leurs données médicales. On les appellera par la suite "**archiveurs**".

La notion d'archive est facilement compréhensible lorsqu'on parle de dossiers médicaux papier, car concrète. Elle devient problématique lorsque transposée dans le domaine informatique où elle devient plus abstraite pour l'utilisateur, a fortiori pour un généraliste.

Pourquoi archiver un dossier papier ? Un dossier qui n'est plus activement utilisé prend physiquement de la place de stockage et ralentit le médecin dans sa recherche d'autres dossiers souvent consultés dans la masse de papier à manipuler. Il y a alors intérêt de le déplacer dans un autre espace de stockage dédié, où les conditions sont propices à la conservation de longue durée du support (papier et encres) et à l'abri du vol.

Selon notre ouvrage technique de référence écrit par Steven NELSON [\(5\)](#), "*Les notions de sauvegarde et d'archivage informatiques sont souvent présentées de façon interchangeables, comme*

une méthode de protection de données qui s'étend sur une certaine période de temps. Le fait que de nombreux logiciels et organisations regroupent ces deux fonctions dans la même catégorie en mettant surtout en avant le versant sauvegarde rend les choses encore plus confuses. Ceci donne l'illusion que ces deux notions ne constituent qu'une seule et unique fonction."

Lors de la conception de notre questionnaire, nous avons considéré l'état des progiciels médicaux concernant leurs capacités d'archivage. En reprenant les travaux du Dr STAECHELÉ qui avait fait un comparatif des fonctions de sauvegarde des principaux progiciels disponibles pour les généralistes, la notion d'archive proprement dite n'est effectivement pas apparente dans les rubriques dédiées à la protection et sauvegarde des données. Il nous a donc été impossible de savoir si la fonctionnalité de sauvegarde de chacun de ces progiciels intégrait par défaut un véritable système d'archivage.

Si la fonction d'archivage au sein des systèmes de sauvegarde progiciels existe sous une forme non explicite, intégrée à la fonction "sauvegarde", les utilisateurs de ces progiciels peuvent ne pas être conscients de son existence. Ceci pourrait expliquer le faible taux d'archiveurs (25%) parmi nos répondants.

Toujours selon Steven NELSON, la notion d'archive informatique est bien distincte de la sauvegarde : *"Dans une base de données, des données archivées ne sont pas des copies itératives de données (comme dans une sauvegarde), mais plutôt des **données** originales qui ont été **déplacées** d'un emplacement à un autre. Ce **déplacement** est décidé **selon des critères assignés aux données**, comme leur ancienneté par exemple. [...] Les archives sont conçues pour stocker ces données sur le long terme, sur un support dont le vieillissement n'induit pas leur dégradation."* [\(4\)](#)

L'auteur explique ensuite que les données à archiver sont dans l'idéal des données **statiques**, pour lesquelles appliquer des méthodes de sauvegarde itératives est un gaspillage de ressources, qui au final n'apporte pas forcément une solution de stockage sur le long terme. Réaliser une archive a donc pour principal objectif de réduire la quantité de données à sauvegarder, et donc le temps nécessaire pour réaliser la sauvegarde. Dans le cas d'un dossier médical, le critère d'archivage le plus évident où les données deviennent statiques est le décès du patient. On pourrait envisager également le changement de médecin traitant au profit d'un autre médecin, ou l'absence d'accès à un dossier pendant une longue période.

L'argument d'épargne des ressources est particulièrement pertinent lorsque les quantités de données sont très importantes, avec des bases de production dépassant le téraoctet (= 1024 giga-octets). Les dossiers patients informatisés chez un généraliste ne présentent pas les mêmes caractéristiques. Le dossier d'un seul patient n'occupe qu'un espace de stockage relativement faible dans la sauvegarde : au plus, plusieurs centaines de méga-octets s'il comprend des images. Dans les résultats du Dr STAECHELÉ en 2012, seuls 42% des répondants avaient une quantité de données médicales informatisées dépassant le giga-octet. En 2018, une base de données progicielle d'un médecin généraliste pèse au plus une à plusieurs dizaines de giga-octets. Quand bien même il y aurait présence de données statiques archivables au sein de la sauvegarde, un ordinateur moderne n'est

pas significativement ralenti lors de la navigation dans une telle base de données. Il pourrait le devenir si la quantité de données à gérer augmentait plus vite que ses capacités de traitement.

La transition du papier au numérique est encore récente. Notre expérience d'exercice sur le terrain, bien que limitée, nous a montré que les généralistes français conservent des archives de dossier papier en parallèle de leur système informatique, et ce pour plusieurs raisons. Numériser l'entièreté d'anciens dossiers papier est chronophage et peu rentable à l'usage. Ensuite, par obligation légale. La législation concernant la durée de conservation des dossiers médicaux n'a pas évolué depuis 2009, et est résumée par l'ordre des médecins comme suit : [\(42\)](#)

“Aucun texte ne fixe pour les médecins libéraux, la durée de conservation de leurs archives. Pour autant qu'ils aient commencé leur activité avant le 5 mars 2002, ils restent soumis à la prescription trentenaire. Si le délai de prescription a été ramené à 10 ans, le point de départ - la consolidation du dommage - fait planer une incertitude sur la durée de conservation. L'alignement sur le délai de 20 ans retenu dans les établissements de santé pour la conservation des archives paraît raisonnable et à recommander.”

Prenons l'exemple de la Bibliothèque Nationale de France : la numérisation de masse est utile pour des documents anciens, rares et précieux [\(19\)](#). L'objectif est de conserver ces documents pour une très longue durée - bien au-delà des 30 ans - et permet de les partager largement ensuite. Les données médicales n'ont pas les mêmes besoins car elles ne sont pas largement partagées et ont, comme les patients qu'elles concernent, une durée de vie limitée.

Ainsi, pour une durée de conservation de 30 ans environ de dossiers informatiques de volume relativement faible, la nécessité d'archiver informatiquement au sens propre du terme n'est pas si évidente pour les généralistes d'aujourd'hui.

Il est difficile d'anticiper l'évolution du rapport entre rapidité de navigation et augmentation de taille des bases de données. D'un côté, une augmentation de la quantité d'espace de stockage nécessaire par patient est prévisible dans les prochaines années : les dossiers papier disparaissent avec la destruction des anciennes archives, et les nouveaux documents créés le sont de plus en plus sous forme numérique. Les médecins ajoutent ces documents numérisés à leurs bases de données en plus de ceux qu'ils peuvent eux-mêmes créer, comme des photos de suivi dermatologique, ou des clichés d'imagerie médicale.

De l'autre, les technologies des supports de stockage s'améliorent avec le temps : leur capacité augmente, et les temps d'accès à l'information qu'ils contiennent se réduisent. Ces augmentations suivent une croissance exponentielle, ce qui n'est pas le cas de la population française.

Une fois que tous les dossiers patients seront numériques, il est possible que la réalisation de véritables archives informatiques devienne plus intéressante voire nécessaire pour conserver des performances d'utilisation du progiciel satisfaisantes.

Pour ce qui est de notre questionnaire, nous avons volontairement conservé cette notion mixte d'archive informatique en simplifiant sa définition en *“un exemplaire de sauvegarde auquel on ne touche plus, stocké à l'abri, pour conservation à long terme.”*.

Parmi les archiveurs, 123 (78,34%) disent ne jamais recopier leurs archives informatiques. Ce chiffre pourrait révéler quelque chose de problématique. Si les données des archives sont existantes en un seul exemplaire, c'est-à-dire effacées des données de travail après leur export, alors ces données sont fragilisées car exposées à une disparition dépendant du temps de rétention de l'information du support d'archivage.

Les supports choisis par la majorité de nos répondants archiveurs ne sont pas idéaux : disque dur standard ou clé USB, les deux ne sont pas suffisamment fiables pour une conservation pendant 30 ans : la pérennité de la majorité des supports physiques est comprise entre 5 et 10 ans. Au-delà de cette durée, la conservation [verbatim](#) de l'information n'est plus assurée. Comme pour la sauvegarde, la problématique de pérennité est primordiale : la solution employée dans l'industrie est de recopier tous les ans les sauvegardes et archives, pour pérenniser l'information qu'elles contiennent. Théoriquement, les meilleurs supports d'archivage longue durée sont les [m-disc](#) ou les bandes magnétiques modernes (sous réserve du respect des conditions de conservation adaptées), mais ce sont des supports coûteux et peu pratiques à manier.

Sur le terrain, un généraliste peut réaliser une purge des dossiers archivés et non consultés depuis 30 ans sans crainte. Puisqu'il est logique d'effacer des données devenues inutiles, une conservation sur support durable unique non recopié n'est pas une méthode utile, pratique ou économique. L'adoption d'un [schéma de sauvegarde incrémentale](#) suivant la [règle des 3-2-1](#) avec sauvegardes régulières peut donc encore aujourd'hui être considéré comme nécessaire et suffisant.

4.9. Score selon qualité de sauvegarde

Le score de qualité de sauvegarde met l'accent sur le processus de restauration pour deux raisons. D'abord, les thèses précédentes sur le sujet ont montré qu'il s'agissait d'un point faible majeur dans les procédures des généralistes : il convenait de ré explorer ce point plusieurs années après. Ensuite, un test de restauration itératif est le seul procédé qui permet de valider le bon fonctionnement d'un système de sauvegarde. C'est également le seul procédé permettant un plan de restauration rapide et efficace, d'où les deux points attribués dans ce cas.

Trois des caractéristiques démographiques de nos répondants sont associées à un meilleur score de sauvegarde de façon statistiquement significative : le sexe masculin, un âge élevé et une ancienneté d'installation élevée.

Il est fréquent de constater des différences entre les hommes et les femmes dans de nombreux travaux. L'explication de cette disparité pourrait être sociétale, avec un plus grand attrait pour l'utilisation des technologies informatiques chez les hommes, mais sort du cadre de notre étude.

Les meilleurs scores des répondants les plus âgés allaient contre l'intuition de meilleurs scores chez les répondants plus jeunes, potentiellement plus habitués au maniement des nouvelles technologies. On considère cependant ici en particulier l'usage d'un progiciel médical : les médecins les plus âgés sont possiblement avantagés par leur expérience. Plus le temps passé à utiliser leur progiciel est long, plus la probabilité d'utilisation poussée et de paramétrage en détail augmente. Également, un exercice plus long peut correspondre à l'emploi de plusieurs progiciels différents, ou de plusieurs versions majeures d'un même progiciel. L'adaptation à un changement de progiciel peut être facilitée par l'expérience : les tâches à effectuer avec un progiciel ne changent pas tellement avec le temps (entrer les données administratives et observations, intégration de fichiers externes, sauvegarde des données...). Les plus grands changements portent sur les méthodes utilisées pour effectuer ces tâches, et sur l'interface avec l'utilisateur. Un réapprentissage pour un progiciel plus récent est facilité par l'expérience gagnée sur un progiciel ancien. Il est aussi possible qu'une longue expérience d'utilisation de progiciel ait exposé nos répondants à davantage d'incidents, et les ait en retour incités à améliorer leurs pratiques.

Les meilleurs scores chez les médecins en exercice en solo étaient significatifs en analyse univariée, mais pas en multivariée. Une puissance plus importante pourrait dégager ce facteur d'étude comme possiblement déterminant dans la qualité du score de sauvegarde, mais nous ne pouvons pas l'affirmer.

4.10. Limites de l'étude

La problématique informatique au cabinet de médecine générale est relativement vaste. Les travaux précédant cette thèse ont montré qu'il n'était pas réaliste d'espérer explorer toute la problématique d'une seule traite. Les aspects non explorés en termes de protection des données ont été, entre autres :

- la protection du poste de travail, physique (verrous, emplacement...) et logicielle (antivirus, pare-feu...),
- l'utilisation de mots de passe,
- les méthodes de transmission des données : boîtes mail sécurisées, VPN, appartenance à un réseau médical agréé.

Nous avons identifié et rencontré plusieurs biais possibles. Premièrement, plusieurs **biais de sélection** : malgré notre intérêt explicitement signalé pour les réponses des médecins généralistes non experts en informatique, nous avons conscience que cette thématique n'intéresse pas tous les médecins au même degré. Le niveau d'intérêt de nos répondants pour ce questionnaire pourrait être en lien avec leur degré d'investissement professionnel dans les technologies informatiques.

Comme pour tout recueil de données où la participation active du généraliste est nécessaire, le temps disponible pour répondre est également un élément sélectionnant possible.

Comme il est expliqué dans la discussion, la médiane d'âge de nos répondants et celle des généralistes français ne correspondent pas, celle de nos répondants était plus basse (entre 41 et 50 ans).

Ensuite, un **biais d'interprétation** pouvait être présent lors des réponses au questionnaire, du fait de l'emploi de concepts et notions informatiques dont le degré de compréhension est variable entre chaque généraliste. Nous avons tenté de réduire ce biais par l'emploi des définitions simplifiées en légende des questions où ces notions apparaissent.

Un **facteur de confusion** est envisageable dans la question « *Selon vous, le médecin généraliste est-il responsable de la sécurité de ses données informatisées : s'il ne fait appel à aucun prestataire ? et s'il en confie la gestion à un prestataire ?* ». **Le problème résidait dans la présentation visuelle des deux items de question.** Proposés ensemble ou séparément ces deux items forcent le répondant à réfléchir à sa réponse. Le répondant pourrait par exemple se demander si les deux cas de figure sont forcément différents.

Enfin, un **biais de désirabilité** est envisageable dans les réponses : pouvoir *a posteriori* comparer ses réponses à celles de tous ses confrères aura pu inciter certains répondants à surestimer leurs méthodes de sauvegarde. Par exemple, on peut supposer que dans certains cas le nombre d'exemplaires de sauvegardes a pu comprendre des sauvegardes partielles ou d'ancienneté différente, qui n'étaient pas à comptabiliser. Nous avons tenté de limiter ce biais en interrogeant des notions aussi factuelles que possibles. Dans le cas des deux questions subjectives (*Quel niveau de confiance accordez-vous à la sécurité de vos données médicales informatisées ?* et *Quel niveau de compétence personnelle pensez-vous avoir concernant la réalisation et la restauration de vos sauvegardes ?*), la collecte des réponses qui suivent permettent de confronter ce ressenti aux faits.

Nous avons également relevé plusieurs **erreurs de conception du questionnaire** au moment de l'analyse des résultats.

Comme il est signalé plus haut, la question portant sur la machine utilisée pour restaurer une sauvegarde était mal programmée et a rendu des résultats aberrants.

Le questionnement sur la vérification des sauvegardes avait une condition d'apparition mal choisie. Elle était de ne pas avoir répondu "Oui" à la question "*votre progiciel est-il en cloud ?*", ce qui équivaut à l'ensemble des réponses "Non" ou "Je ne sais pas" à cette question. Elle aurait plutôt dû être posée à tous ceux ayant réalisé au moins une sauvegarde physique. En l'état, elle exclut tous les répondants avec progiciel *cloud*, y compris ceux réalisant des sauvegardes sur supports physiques en plus. On sait cependant grâce à la question sur les supports physiques que sur les 106 répondants en progiciel *cloud*, 72 n'utilisent aucun support physique en complément. On a donc perdu l'opportunité d'une réponse pour 34 participants sur ce point.

La fréquence de réalisation des sauvegardes n'a pas été questionnée. Ce manque n'a pas été remarqué malgré plusieurs revues et bêta-tests du questionnaire. On suppose que cette erreur de conception n'a pas été remarquée car nous interrogeons la fréquence d'archivage en fin de questionnaire. Heureusement, les résultats portant sur les pertes de données avec des pertes

définitives peuvent nous aider à compenser cette erreur. Puisque les pertes de données définitives sont majoritairement inférieures à une semaine, on peut supposer que la fréquence de sauvegarde la plus répandue est inférieure à une semaine, même si nous ne pouvons l'affirmer.

4.11. Points forts de l'étude

Notre travail a pour lui son **originalité** : le thème de l'informatique en médecine générale n'a pas été souvent traité.

Par rapport aux travaux précédents, grâce à une **diffusion** sur une zone géographique **plus large**, le **nombre de réponses au questionnaire obtenu** est bien **supérieur**. Le nombre relativement important de répondants nous a permis d'obtenir des **résultats statistiquement significatifs**. Le questionnaire aurait pu être encore plus largement diffusé, par exemple via des groupes de médecins sur les réseaux sociaux, mais il aurait perdu de sa validité : une diffusion via les URPS garantit que les répondants contactés sont effectivement médecins généralistes en exercice.

Les **méthodes** de mise en place du questionnaire sont **inhabituels et détaillés**, et pourraient servir à de futurs collègues thésards qui souhaiteraient employer des méthodes d'auto-hébergement de questionnaire en ligne.

Les éléments de notre critère composite sont objectifs, et laissent peu de place à une mauvaise auto-évaluation par le répondant.

L'investissement en matériel informatique fait par le thésard était faible car le matériel employé était peu coûteux ou déjà disponible. Le coût d'un hébergement de questionnaire en ligne a été évité. L'avantage financier a par contre été largement contrebalancé par le temps nécessaire à la mise en place des serveurs de questionnaire.

4.12. Réflexions à l'issue du travail

Nous nous étions lancés dans ce travail avec enthousiasme, confiants de nos compétences pour réaliser un questionnaire en ligne. Nous avons rencontré bon nombre de difficultés, et avons eu besoin de presque deux ans pour achever ce travail. Plusieurs aides techniques nous ont été nécessaires et précieuses. Nous avons donc réalisé que nous étions sujets à une discordance confiance-compétence tout autant que nos répondants. Ainsi, il nous semble nécessaire que les généralistes s'entourent de soutiens compétents dans ce domaine, même s'ils sont persuadés de bien faire.

L'adoption large des nouvelles technologies de sauvegarde est limitée par la nature concurrentielle et fragmentée du marché des progiciels. Contrairement au monde médical où de grandes institutions comme l'HAS ou les collèges de professeurs émettent des recommandations consensuelles, il n'existe pas en informatique de "bonnes pratiques" établies. Les pratiques de chaque éditeur de logiciel sont régies par trois facteurs principaux : la loi, leur cahier des charges et les besoins de leurs

clients. Un progiciel peut ne pas offrir par défaut des fonctions perçues comme indispensables comme la sauvegarde. Si le progiciel respecte la loi et le cahier des charges mais que l'éditeur n'est pas enclin à intégrer une nouvelle fonction, ou réparer une fonction problématique, il peut être nécessaire d'en changer. La concurrence peut être une chance pour le consommateur car elle peut entraîner réduction des prix et avancées techniques pour viser à se démarquer, ce qu'ont fait les progiciels en cloud. Cependant, un généraliste peut être découragé de changer son progiciel à cause de difficultés d'export de ses données d'un progiciel à l'autre.

Dans le monde de l'informatique, il n'existe pas de format de fichier universel pour un type d'information donnée. Chaque méthode de numérisation a ses avantages et ses inconvénients, et certains formats sont propriétaires : un logiciel X ne peut souvent pas utiliser les fichiers d'export spécifiques du logiciel Y. On peut parfois convertir les données du format X à Y, mais l'information finale est trop souvent altérée lors du processus de conversion.

Pour les généralistes, il n'existe pas non plus de format universel d'export des données de progiciels. L'information peut être exportée sous différentes formes, mais nécessitera forcément un travail de réorganisation lors de l'import si le progiciel de destination est différent du progiciel source. En pratique, les formats les plus utilisés sont des formats d'usage général : les fichiers texte ou PDF. L'information sera présente, mais perd sa catégorisation en champs de base de données. On perd donc du temps à sélectionner des fragments d'information pour les placer aux endroits adéquats, et s'agissant d'une méthode manuelle il peut y avoir des erreurs ou oublis. Les progiciels les plus récents proposent des systèmes d'importation depuis les formats concurrents, mais ce n'est pas une norme dans le secteur.

Les outils disponibles aujourd'hui devraient assurer une meilleure qualité de sauvegarde que celle qui a été constatée. Les services rendus par les technologies informatiques évoluent constamment, ainsi que les connaissances à avoir pour les employer. Ces connaissances ne relèvent cependant pas de l'exercice médical proprement dit. Il est encore possible de sauvegarder correctement ses données médicales, au prix d'un investissement que tous les généralistes risquent de ne pas faire. Si un généraliste ne souhaite pas s'investir, il lui faut déléguer les tâches informatiques à sa charge, dont la sauvegarde, à un tiers certifié.

On se propose donc à l'issue de ce travail de participer à l'effort de prise de conscience en réalisant une fiche de synthèse. Elle mentionne les points perfectibles des méthodes de sauvegarde à corriger ou faire corriger. Elle sera envoyée à tous les URPS ayant participé à la diffusion de notre questionnaire, et pourra être diffusée en congrès.

A notre avis, l'amélioration des pratiques de sauvegarde passera en grande partie par l'évolution progicielle, dont le fer de lance actuel est le progiciel cloud. L'adoption des nouvelles méthodes de sauvegarde ira de pair avec l'intégration d'autres avancées : il pourrait y avoir un partage de l'information plus facile entre deux utilisateurs du même progiciel. L'information médicale ne circulerait plus entre les ordinateurs des médecins : il y aurait seulement attribution de droits d'accès

d'un médecin à un autre, l'information restant physiquement centralisée. Il s'ensuivrait un gain de temps potentiel par l'absence de transfert ou de conversion de données. De plus, un système centralisé avec un classement d'information standardisé permet la recherche clinique sur des données de terrain, comme vu dans [l'exemple britannique de la UK General Practice Research Database](#). Malheureusement, le marché du progiciel en cloud est aussi fragmenté, limitant ces bénéfices potentiels. Pour améliorer le service rendu au patient tout en respectant ses droits et sa vie privée, il est souhaitable que l'information médicale soit plus largement partagée entre praticiens de santé, mais de façon sécurisée et confidentielle. Une sauvegarde de qualité fait partie des piliers de ce progrès.

Les généralistes resteront par leur travail les premiers consignataires de l'information médicale. Le patient serait suivi par ses informations médicales et n'aurait pas à faire saisir de nouveau les mêmes informations dans un autre système. Le médecin n'aura lui pas à exporter ses données médicales dans un autre format, ce qui on l'a vu est chronophage et peut faire perdre de l'information.

Un exemple notable français d'effort de regroupement national de l'information médicale standardisée avec des mesures de sécurité adéquates est le DMP ou Dossier Médical Partagé (connu avant 2015 sous le nom de Dossier Médical Personnel). Cet outil a longtemps été sous-exploité par les généralistes. Son image est très écornée par son histoire laborieuse de développement à grand frais, et ses résultats concrets médiocres : utilisation chronophage, ergonomie insuffisante, problèmes techniques, mauvaise intégration aux autres outils disponibles. En sus, la sécurité et confidentialité des données hébergées n'étaient pas solidement garanties. En 2018, une refonte du système DMP a été réalisée. Ce **nouveau DMP est encadré par la législation des Hébergeurs agréés de Données de Santé**. Il est donc désormais plus à même d'assurer la sécurité et la confidentialité des données qui lui sont confiées que la plupart des généralistes. Malgré la refonte de 2018, l'emploi du DMP reste malaisé. De nombreux médecins qui s'essayaient à l'exercice sont rebutés par ce nouveau DMP encore incompatible avec la réalité du terrain [\(43\)](#).

L'emploi par les généralistes des avancées informatiques au sein des progiciels tarde, car leur intégration dépend des éditeurs. Mais l'emploi de progiciels aux méthodes modernes est aussi freiné par d'autres facteurs. En médecine générale, faire preuve de bon sens, patience et prudence dans la prise de décision est nécessaire. Raisonner de cette façon par rapport à l'informatique empêche de se positionner à la pointe de la technologie informatique, et peut inciter le généraliste à se reposer sur des méthodes testées et approuvées, évitant les déboires de systèmes récents mais non encore éprouvés.

Quelques-uns de nos répondants étaient clairement au courant des dernières avancées. Au début de notre travail, la certification HDS de 2018, n'existait pas encore : appliquée en avril, elle n'était pas intégrée dans le détail de nos questions. Cela n'a pas empêché un de nos répondants de nous

demander par mail pourquoi le cas de figure “*cloud* tiers agréé HDS” n’avait pas été proposé. Le questionnaire avait débuté en juillet : ce répondant était donc au courant des dernières évolutions après la modification de la législation. C’est un témoignage de l’investissement d’une partie de la population médicale sur le terrain informatique. La grande majorité des répondants n’était cependant pas dans cette situation.

L’inconvénient de l’hébergement *cloud* ou décentralisé chez un HDS est de désigner clairement une cible : un HDS serait dépositaire d’une grande quantité d’informations sensibles. Pirater ces données serait sans doute plus difficile que celles d’un généraliste lambda en particulier, mais le butin potentiel en serait bien plus important, et donc plus attrayant.

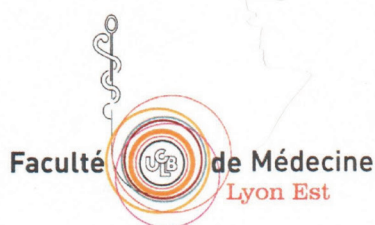
De plus, la perception du label HDS pourrait être affectée par sa démocratisation : plusieurs entreprises privées sont désormais certifiées HDS, comme la maison-mère de l’opérateur de téléphonie Free, Iliad [\(44\)](#) et le géant informatique Microsoft [\(45\)](#). Puisque ce sont des normes ISO préexistantes qui servent désormais de socle à la certification, il est devenu beaucoup plus simple d’obtenir le droit d’héberger des données de santé. Les opérateurs privés prennent la main sur un nouveau marché juteux spécifiquement français, ce qui pourrait brouiller la confiance à apporter à ce label. Un généraliste pourrait penser comme suit : confier ses données à une entreprise privée sans lien direct avec le soin, n’est-ce pas potentiellement risqué quand l’objectif du privé peut être la monétisation des données de santé ? Quid de la confidentialité confiée à ces hébergeurs ? Ce sont pourtant ces aspects que vise à encadrer le label HDS.

Quels que soient les moyens choisis par les généralistes pour renforcer la protection et les sauvegardes de leurs données médicales informatisées, il conviendra de rester critique par rapport aux méthodes facilitantes proposées.

Il y a matière à réalisation d’un autre travail de thèse, qualitatif cette fois, pour explorer les raisons de ce **manque de confiance numérique français**, et les **freins des généralistes à l’utilisation du nouveau DMP**. Les freins pressentis étant :

- la méconnaissance des HDS et de la charte d’utilisation,
- le manque de confiance numérique en les institutions de santé gouvernementales,
- la peur de piratage d’une base de données centralisée [\(46\)](#).

5. Conclusions



Nom, prénom du candidat : MAGUY Paul

CONCLUSIONS

Le RGPD a contribué à redonner de l'importance aux méthodes de protection des données médicales. Nous avons réalisé en 2017-2018 une étude quantitative concernant les connaissances et pratiques des médecins généralistes français. Notre questionnaire en ligne diffusé par sept URPS a recueilli 609 réponses.

Parmi les médecins répondants, 98% réalisaient des sauvegardes. Au cours des 5 dernières années, 18% avaient subi une perte de données, au moins en partie définitive pour la moitié d'entre eux. Seuls 25% des médecins ont déjà réalisé un test de restauration (unique pour 6% des médecins).

La sécurité des données médicales informatisées suppose une implication et une prise de conscience des généralistes. À défaut d'appliquer des corrections eux-mêmes à leur système informatique, ils doivent s'assurer que leur progiciel utilise des méthodes adaptées.

Pour limiter les risques de perte de données dues à des mises à jour de logiciels ou à la rupture d'une chaîne d'opérations, il est important de regrouper les données médicales au sein du progiciel et d'utiliser l'outil de sauvegarde intégré au progiciel.

Pour limiter les risques de perte de données entre deux sauvegardes, une automatisation du processus de sauvegarde est souhaitable.

Les méthodes de chiffrement des données sont mal connues des généralistes. Toutes les données médicales destinées à l'export à un tiers doivent être chiffrées. Elles devraient être chiffrées par défaut par les progiciels.

Dans le cas d'un progiciel en *cloud*, il faudrait réaliser annuellement une exportation des données du *cloud* sur un support local, en plus des vérifications faites par le fournisseur de service.

En cas de difficulté de gestion de son système de sauvegarde local, ou de volonté d'héberger ses données médicales chez un tiers en plus de ses propres sauvegardes, il est conseillé de se diriger vers un éditeur de progiciel agréé Hébergeur de Données de Santé. L'agrément HDS autorise une redistribution des tâches informatiques et libère du temps pour le généraliste.



La réalisation spécifique d'archives informatiques n'est ni obligatoire ni nécessaire. Cependant, il est obligatoire d'avoir un système de sauvegarde fiable englobant l'intégralité des données médicales et de respecter les délais de conservation conseillés par l'Ordre des Médecins.

Vérifier la réussite du processus de sauvegarde permet de repérer une erreur critique corrigible. Pouvoir paramétrer le logiciel de sauvegarde pour afficher automatiquement un message en cas d'erreur du processus de sauvegarde est important.

Le Président de la thèse,
Nom et Prénom du Président
Signature

Le Président de la thèse

Vu :
Pour le Président de l'Université,
Le Doyen de l'UFR de Médecine Lyon Est



Professeur Gilles RODE

Vu et permis d'imprimer
Lyon, le **03 AVR. 2019**

6. ANNEXES

6.1. Fiche de bonnes pratiques

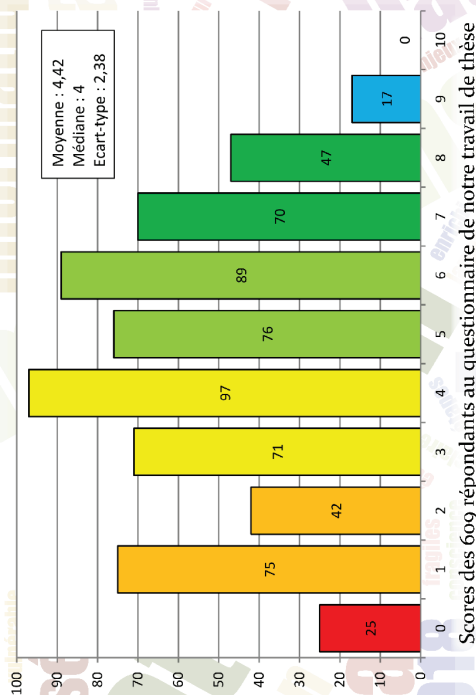
Médecins Généralistes : Évaluez la qualité de sauvegarde de vos données médicales informatisées !

Dans le cadre de notre travail de thèse intitulé « *Protection, sauvegarde et archivage des données médicales informatisées par les médecins généralistes français en 2017-2018* », nous avons dégagé les critères de qualité de la sauvegarde de données médicales informatisées d'un médecin généraliste.

Ils constituent un **score de qualité de sauvegarde**, entre 0 et 10 points :

- Votre sauvegarde devrait :
 - contenir l'intégralité des données médicales : } 1 point
 - toutes les données du progiciel
 - et les données médicales hors progiciel,
 - respecter les éléments de la **règle des 3-2-1** :
 - 3 exemplaires des données de travail*1 point
 - sur 2 types de supports différents,.....1 point
 - dont 1 copie hors site,.....1 point
 - être automatisée et quotidienne,.....1 point
 - utiliser l'outil de sauvegarde intégré au progiciel,.....1 point
 - être protégée par un chiffrement informatique.....1 point
- Chaque généraliste devrait :
- vérifier la réussite du processus de sauvegarde,.....1 point
 - réaliser un test de restauration :
 - au moins une fois.....1 point
 - ou au mieux
 - plus d'une fois (tous les ans).....2 points

*Comprend l'exemplaire de travail = les données utilisées au quotidien



Et vous, quel est votre score de sauvegarde ?

Si votre score n'est pas d'au moins 8/10, vous devriez contacter votre fournisseur de progiciel pour faire appliquer des mesures correctives.

En cas de difficulté de gestion de votre système de sauvegarde, ou si vous envisagez de changer de progiciel, nous vous conseillons de vous diriger vers un éditeur de progiciel agréé **Hébergeur de Données de Santé**. Cet agrément permet de déléguer en confiance des tâches informatiques, ce qui vous libère du temps de soin.

Pour en savoir plus, consultez notre travail de thèse disponible sur Internet via le **SUDOC** (<http://www.sudoc.abes.fr>).

MAGUY Paul - Dr LAPRAIS Olivier

6.2. Résumé détaillé des travaux de thèse précédents

On résume ici les thèses précédentes sur le thème de l'informatique, en ne conservant que les points en rapport avec notre travail. Nos remarques quant à ces travaux sont en italique.

- 2009 - Dr BRAMI Gregory (PARIS 7)

Protection des données patient informatisées en médecine générale Étude transversale 11/2007 - Département du Val d'Oise

Il s'agissait d'une étude transversale adressée par questionnaire à 280 médecins du Val d'Oise entre septembre et décembre 2007. 149 médecins avaient répondu au questionnaire.

Résultats : **les médecins interrogés possédaient des connaissances en sécurité informatiques**

- utilisation d'un mot de passe (79%)
- utilisation d'un antivirus (78%) et sa mise à jour (80%)
- élaboration de sauvegardes (99%)
- copie des sauvegardes (95%)

Ces connaissances avaient leurs limites et beaucoup de mesures n'étaient que peu voire pas utilisées pour assurer la sécurité et la pérennité nécessaires aux données médicales :

- utilisation de mots de passe forts (1%)
- renouvellement du mot de passe (4%)
- vérification de la « lisibilité » des sauvegardes (40%)
- conservation de copies de sauvegarde hors du cabinet (67%)
- utilisation d'autres logiciels de protection, la protection contre le vol (64%)

Or, **la perte de données et le risque informatique mesurés étaient réels** :

- 31% des médecins interrogés ont perdu des données
- 24% ont été victimes d'un virus informatique.

Et malgré ces chiffres, 63% avaient confiance dans la protection de leur système informatique.

La conclusion de ce travail a été l'élaboration d'une **liste de mesures simples pour assurer une protection minimale contre la perte de données**, dont :

- Faire des **sauvegardes régulières**.
- Faire des **copies de sauvegarde sur différents supports**.
- **Vérifier régulièrement la validité des sauvegardes** effectuées.
- Conserver des **copies de sauvegarde en dehors du cabinet**.
- Utiliser des **logiciels de protection** (au moins un antivirus) et **les mettre à jour régulièrement**.
- **Protéger son ordinateur** contre la **détérioration criminelle ou accidentelle**.

- 2012 - Dr CHERIET Mondher (PARIS 6)

Sécurité des données médicales informatiques en cabinet de médecine générale

Il s'agissait d'une étude conduite en 2011. 165 questionnaires avaient été recueillis auprès de médecins généralistes français parmi les 2 380 sollicités. Les questionnaires avaient été envoyés via la mailing-list de la SFTG (Société de Formation Thérapeutique du Généraliste).

Les résultats de l'étude ont été classés selon des critères composites pour le Réseau, la Sauvegarde, les Mots de passe, les Logiciels de protection, la Maintenance et le Dossier Médical Personnel; avec un classement en A, B et C en fonction du degré de conformation aux mesures sécuritaires, A étant la meilleure classe.

- **27,8% déclaraient avoir été victime d'au moins une entrée par effraction concernant leur cabinet**, dans un tiers des cas accompagné du vol de matériel informatique, <5% disque dur seul, 30% ordinateur complet, 65% pas de vol informatique
- pour la classe "sauvegarde" :
 - **un tiers était en classe A (une ou plusieurs sauvegarde, plus d'une fois par semaine, hors cabinet médical),**
 - **2/3 en B (idem mais une fois par semaine ou moins, ou si plus fréquent uniquement au cabinet),**
 - **<5% en C (aucune sauvegarde).**
- Comparaison du résultat obtenu de 41% de sauvegarde quotidienne versus une étude de 2007 dans une revue informatique américaine à 77%.
- **La qualité des sauvegardes était non garantie pour 69% des répondants** car sauvegarde jamais testée, 47% n'étaient pas sûr que leur sauvegarde contenait tous les fichiers importants.

Les conseils qui suivaient ces résultats étaient :

- Augmentation de la fréquence de sauvegarde à 2/semaine
- Conservation des sauvegardes à l'extérieur du cabinet
- Utilisation de sauvegarde incrémentale (*permet la conservation d'anciens fichiers écrasés sans [rotation des supports](#) de sauvegarde type Grandfather- Father-Son(47).*)

Le Dr CHERIET imaginait alors les progrès possibles pour les généralistes :

- Evocation des solutions de type *Cloud*, *mais à cette époque existait un vide juridique sur ces méthodes concernant les données de santé. Depuis existe la certification des Hébergeurs agréés de données de santé à caractère personnel.*
- Une utilisation distante de toutes les données, plus simplement des sauvegardes incrémentales automatisées distantes, sous la forme d'images disque codées, inexploitable sans mots de passe (*mieux : chiffrement*).

- 2012 - Dr STAECHELÉ Roland (STRASBOURG)

Sécurité des données informatiques

À propos de la sauvegarde des données de santé au cabinet de médecine générale

Il s'agissait d'une enquête qualitative, menée dans 31 cabinets de médecine générale du Haut-Rhin en 2012, avec entretien libre puis directif. Malgré la méthode qualitative employée dans cette étude, il s'agit du travail le plus poussé sur le concept de sauvegarde et d'application pratique au cabinet médical.

Pertes de données

45% des utilisateurs avaient déjà subi des pertes de données. Le temps de travail perdu était à 72% inférieur à une semaine. **Perte irrémédiable dans 2 cas/31 (14%).**

L'origine matérielle était surestimée tandis que l'origine humaine était sous-estimée. Erreurs de manipulations nombreuses, surtout en multi-utilisateurs, et remarquées tardivement lorsqu'un problème apparaît.

Le vol de support de données apparaissait comme un facteur en hausse, devant être pris en compte dans la sécurité locale du cabinet et dans la gestion des supports de sauvegarde (externalisation).

Responsabilité de la sécurité des données

Près de 1 utilisateur sur 3 pensait ne pas être responsable légalement de la sécurité des données de santé qu'il héberge, alors que les données de santé sont considérées au regard de la loi comme des informations sensibles.

Pratiques de la sauvegarde

Sur les 31, 4 (13%) n'avaient jamais fait de sauvegarde et risquaient la perte d'années de travail ! Plus de 50% citaient des difficultés à la gestion d'un système de sauvegarde.

Il y avait appel à prestataire pour paramétrage et gestion des sauvegardes par 40% des utilisateurs, d'où une sous-traitance partielle ou totale, ce qui pouvait provoquer des problèmes de procédures liés à gestion parcellaire et manque d'autonomie de l'utilisateur.

Dans le cas d'appel à prestataire, trois modes de fonctionnement se dégagèrent, classés par ordre de délégation d'autonomie croissante :

- paramétrage initial par le prestataire avec création de chemin de sauvegarde type. L'utilisateur avait accès aux logiciels utilisés, était libre de modifier les réglages mais généralement s'en tenait aux actions prédéfinies : lancement d'une application et connexion au support.
- le prestataire gérait l'essentiel du système par télémaintenance : l'utilisateur ne pouvait pas faire de modification, les routines de sauvegarde étaient des scripts à exécution automatique. La seule action locale à réaliser était la permutation des supports de sauvegarde.
- le prestataire gérait l'ensemble du système de sauvegarde par télémaintenance. Les données étaient stockées vers un support Internet. L'utilisateur ne pouvait apporter aucune modification, et aucune action locale n'était nécessaire.

Double gérance du système : responsabilité en cas de problème ?

Il avait été remarqué la possibilité pour un utilisateur non autonome de déclencher des sauvegardes inefficaces sans avoir en retour un message d'erreur : **sur les 31 cabinets, 3 avaient une**

sauvegarde inefficace.

Il se posait également un problème de sécurité potentiel par l'utilisation de télémaintenance quand la méthode d'accès aux ordinateurs du cabinet médical était peu à pas sécurisée.

Propositions :

- pratiquer de façon impérative une sauvegarde régulière,
- utiliser les logiciels de sauvegardes simples intégrés au progiciel et à l'OS,
- demander au prestataire une aide pour le paramétrage et l'ergonomie,
- garder l'autonomie complète et la maîtrise de ce système de sauvegarde,
- **surveiller et vérifier régulièrement les historiques de sauvegarde.**

Plan de sauvegarde

Données de santé

Moins d'un utilisateur sur 2 (48%) connaissait l'emplacement des données de santé sur le réseau ou sur un poste ! Dans un logiciel de sauvegarde, la sélection de ces données pouvait donc se révéler fastidieuse voire impossible.

Problème de la sélection des données de santé

Installation progiciel monoposte

Selon le progiciel, les répertoires d'installation des fichiers sont complexes à mettre en évidence (répertoires systèmes, cachés, pas facile d'accès pour utilisateur lambda). Dans ces répertoires, les fichiers de Base de Données (BD) sont souvent stockés sous forme brute, sans icône ni extension et uniquement manipulables via le Système de Gestion de Base de Données (SGBD).

La sélection des données de santé dans un logiciel classique est quasi impossible. L'utilisateur peut sélectionner l'ensemble du disque dur, mais le problème se déplacera lors d'une tentative de restauration. Par ailleurs, un logiciel classique peut ne pas prendre en compte les traitements spécifiques à la BD utilisée, et les fichiers de santé sauvegardés risquent d'être corrompus ou illisibles.

La solution reste simplement d'utiliser l'outil de sauvegarde intégré au progiciel. Les fichiers de santé sont déjà listés dans l'algorithme de traitement du logiciel. La sélection est automatique pour la sauvegarde et la restauration. Par ailleurs, les traitements spécifiques à la BD utilisée permettent de sauvegarder les fichiers dans un état cohérent.

Données hors progiciel

Plus d'un tiers (36%) des utilisateurs sauvegardaient des données non gérées par le progiciel médical : données de santé (messagerie Apicrypt ou professionnelle, carnet d'adresses, agenda, gestion de Rdv, etc.) ou personnelles (iconographies, rapports, etc.).

Proposition de schéma si données hors progiciel

La sélection de ces données est effectuée avec un logiciel de sauvegarde classique en complément de l'outil de sauvegarde progiciel.

La sauvegarde progicielle serait faite dans un répertoire interne au poste de travail, puis sauvegarde classique qui reprendra ce répertoire et les données hors progiciel vers un support externe. Cette méthode permettrait de simplifier la procédure progicielle et de réserver la gestion des supports externes à la sauvegarde classique.

Les logiciels de sauvegarde système (Windows ou MacOS) proposent des assistants au paramétrage. Habituellement tout le profil utilisateur est présélectionné y incluant les répertoires plus difficiles à localiser (messagerie, contacts...).

Supports de sauvegarde

Les supports externes (clés USB, disques durs) sont largement utilisés.

AUCUN cabinet n'utilisait le cryptage matériel proposé sur ce type de support.

Un cabinet sous-traitait entièrement la sauvegarde vers un hébergeur non agréé ASIP.

Le choix du support de sauvegarde tient compte de plusieurs facteurs :

- la distribution des données dans le réseau,
- le volume total des données,
- la transportabilité et la possibilité d'externalisation du support.

Au vu du faible volume de données d'un cabinet médical, on favorisera un support externalisable, type disque dur format poche. Ce support dispose de différents avantages :

- faible coût donc multipliable,
- capacité de stockage importante,
- moins de risque de perte que des clés USB,
- encombrement réduit, facile à externaliser,
- cryptage matériel pour protéger le contenu,
- connectique simple (un seul câble),
- possibilité de dissimuler le support avec un câble long (<5m).

Propositions :

- **utiliser des disques durs de poche comme supports de sauvegarde serveur ou monoposte,**
- protéger les données à l'aide du **cryptage matériel,**
- dissimuler le support en place avec un câble long,
- **externaliser les supports,**
- si besoin, concentrer la sauvegarde des postes clients vers le serveur en utilisant un point de partage.

Stratégie de la sauvegarde

un tiers des utilisateurs effectuaient la sauvegarde avec une fréquence trop faible (hebdomadaire) et disposaient d'un seul support.

En cas de panne, la restauration dépendra de la lisibilité de ce support, ainsi que des fichiers de sauvegarde. Si la restauration se passe sans problème, le temps de travail perdu sera inférieur ou égal à une semaine.

Plus de la moitié (47%) des utilisateurs n'externalisaient pas le(s) support(s).

En cas de vol au cabinet, du serveur et supports annexes, la perte des données de santé sera irréversible, avec des conséquences très importantes sur le travail quotidien et le suivi des patients.

Plus de 50% des utilisateurs ne disposent pas de sauvegardes supérieures à un mois.

En cas d'erreur reconnue tardivement, ou de volonté de retrouver une ancienne version d'un fichier, il sera impossible d'avoir un historique supérieur au mois.

La stratégie de sauvegarde vise plusieurs objectifs :

- **multiplier les supports de sauvegarde,**
- pratiquer une **sauvegarde alternée sur ces supports,**
- **externaliser un ou plusieurs supports,**
- **disposer de sauvegardes les plus récentes possibles,** en heures ou en jour,

- **disposer de sauvegardes anciennes**, en mois ou plusieurs mois.

Plusieurs schémas de rotation des supports/sauvegardes remplissent ces objectifs. Le plus connu est le schéma générationnel GPF (Grand-père, père, fils).

On attribue un jeu de supports à chaque génération et l'on pratique une rotation des supports dans les jeux grand-père/père. Habituellement, les trois générations sont représentées par les sauvegardes quotidiennes, hebdomadaires et mensuelles.

Propositions :

- **Sauvegarde quotidienne** sur deux disques durs permutés tous les lundis. Garder un support externalisé. Via le logiciel, maintenir une rétention de sauvegardes de 15 jours sur chaque support.
- **Sauvegarde mensuelle** sur un disque dur tous les derniers vendredis du mois. Garder ce support externalisé. Via le logiciel, maintenir une rétention de sauvegardes de 12 mois.

Procédures logicielles

La moitié des utilisateurs sauvegardent les fichiers progiciels avec une méthode non supportée par l'éditeur !

La majorité des utilisateurs pensent que la sauvegarde des BD progicielles ne nécessitent pas de procédures particulières.

Si besoin, la sauvegarde classique n'est utilisée qu'en complément de la sauvegarde progicielle.

Propositions :

- **Utiliser exclusivement l'outil progiciel pour la sauvegarde des données progicielles**
- **Doubler avec une sauvegarde classique pour d'autres types de données ou pour disposer des paramètres avancés**
- **Automatiser la sauvegarde classique et progicielle si possible.**

Validation du plan de sauvegarde

Contrôle des procédures

La majorité (68%) ne pratique aucun contrôle des procédures de sauvegarde.

Problématique de la dégradation des procédures

Les processus informatiques subissent fréquemment une dégradation avec le temps, et subissent la loi de Murphy "si ça peut tourner mal, ça tournera mal"

Quelle qu'en soit la source, un dysfonctionnement doit être détecté le plus tôt possible pour pouvoir y remédier.

Propositions :

- contrôle des fenêtres et messages affichés lors d'une sauvegarde manuelle
- surveillance régulière de l'historique des opérations de sauvegarde
- vérification de l'existence, de la date et taille des fichiers produits
- contrôle ponctuel des supports de sauvegarde
- gérance unique des systèmes de sauvegarde

Test de restauration

La majorité (72%) n'a JAMAIS testé la procédure de restauration des fichiers.

Propositions :

- prouver l'efficacité du système de sauvegarde par la restauration

- effectuer régulièrement des tests de restauration des données progiciels sur serveur ou machine hors système.
- Disposer d'une machine hors système pour tester et mémoriser la procédure de restauration.

Conclusions :

- perte d'autonomie de l'opérateur et sous-traitance : **le médecin doit avoir l'autonomie et la maîtrise totale de son système de sauvegarde.** Si besoin, le prestataire pratique un transfert de compétence pour laisser au médecin la gestion complète du système.
- Respect des procédures logicielles : les données de santé sont organisées en base de données qui ne peuvent être sauvegardées de façon sécurisante avec un logiciel classique. **L'outil de sauvegarde fourni avec le progiciel est le seul qui puisse garantir la sauvegarde des bases dans un état cohérent.** C'est aussi le seul outil qui permette la sélection automatique des fichiers progiciels lors de la sauvegarde et lors de la restauration.
- Gestion des supports : **la rotation des sauvegardes sur plusieurs supports sécurisés (cryptage) et externalisables (disque dur externe) est à conseiller.**
- **Surveillance du système de sauvegarde** : la surveillance régulière des historiques et des opérations de sauvegarde est indispensable pour détecter les erreurs et pratiquer une gestion proactive
- **Test de restauration des données** : la fiabilité de toutes les étapes de la chaîne de sauvegarde doit être démontrée. Seule la restauration des données qui permet de retrouver un progiciel fonctionnel peut apporter la preuve de l'efficacité du système de sauvegarde. Un test de restauration peut être pratiqué régulièrement sur le serveur en cours ou une machine de test.

6.3. Label ASIP e-santé

Extrait du document de l'ASIP : "Le référentiel fonctionnel du label « e-santé Logiciel Maisons et Centres de santé » précisant les fonctionnalités attendues en cible pour les niveaux Standard et Avancé de labellisation" [\(48\)](#). On ne présente ici que les items se rapportant à la sauvegarde, archivage et sécurité du progiciel agréé.

Tableau 55 : Référentiel fonctionnel label e-santé Logiciel Maisons et Centres de Santé

	Gestion de la structure	
	Gestion des données	
	Archivage	Archiver le contenu d'un dossier
PILS.1.1.1	Archivage de tout ou partie d'un dossier patient (données administratives et médicales)	<u>Le système DOIT permettre d'archiver tout ou partie d'un dossier patient (données administratives et médicales)</u>
PILS.1.1.2	Fonction de consultation des archives	<u>Le système DOIT permettre de désarchiver tout ou partie d'un dossier patient (données administratives et médicales)</u>
	Paramétrage des sauvegardes	Paramétrer les sauvegardes des données du système
PILS.1.2.1	Paramétrage des sauvegardes automatiques	<u>Le système DEVRAIT permettre de maintenir le paramétrage des sauvegardes automatiques de l'ensemble de la base de données</u>
	Sécurité du SI	
	Gestion des autorisations	
	Autorisations	Gérer les autorisations d'accès au système
SEC.1.1.1	Gérer les autorisations des utilisateurs	<u>Le système DOIT permettre de gérer les autorisations octroyées à un utilisateur en fonction de</u>

		<u>son identité et de son rôle, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.1.1.2	Gérer les rôles	<u>Le système DOIT permettre de gérer des rôles pour l'octroi d'autorisations, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.1.1.3	Gérer l'historique des modifications du paramétrage d'une entité	<u>Le système DEVRAIT gérer un historique de toutes les modifications de paramétrage d'une entité, une règle, un événement du système.</u>
Identification et authentification des acteurs		
	Authentification d'identité	Authentifier les utilisateurs ou les entités avant de leur donner accès au système
SEC.2.1.1	Authentifier les utilisateurs accédant aux ressources protégées du système	<u>Le système DOIT authentifier les utilisateurs accédant au système, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.2	Contrôle d'accès par carte de la famille CPx pour tous les utilisateurs en détenant une	<u>Le système DOIT contrôler l'accès au SI par carte de la famille CPx pour tous les utilisateurs en détenant une (avec le cas échéant rapprochement dans la solution de l'identité de l'utilisateur pour les cartes non directement nominatives, via une fonctionnalité d'enregistrement des cartes de la famille CPx dans le système)</u>
SEC.2.1.3	Contrôle d'accès par "Login / Password" pour les autres utilisateurs	<u>Le système DOIT contrôler l'accès au SI via d'autres méthodes d'authentification documentée (ex : OTP, "Login / Password", CPx d'accompagnement,...) pour les utilisateurs non porteurs de CPx conformément aux pratiques du secteur</u>
SEC.2.1.4	Gérer les données d'authentification de manière sécurisée	<u>Le système DOIT gérer les données d'authentification de manière sécurisée</u>
SEC.2.1.5	Maintenir des règles paramétrables de protection contre les tentatives d'usurpation d'identité	<u>Le système DOIT maintenir des règles paramétrables de protection contre les tentatives infructueuses d'authentification, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.6	Maintenir les délais de péremption des mots de passe	<u>SI les mots de passe sont utilisés ALORS le système DOIT permettre de maintenir des délais de péremption paramétrables pour les mots de passe, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.7	Maintenir la limite de réutilisation des mots de passe anciens	<u>SI les mots de passe sont utilisés ALORS le système DOIT permettre de maintenir une limite paramétrable de réutilisation des mots de passe récents, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.8	Maintenir les règles de sécurisation des mots de passe	<u>SI les mots de passe sont utilisés ALORS le système DOIT permettre de maintenir les règles de sécurisation des mots de passe, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.9	Capturer les mots de passe en les masquant	<u>SI les mots de passe sont utilisés ALORS le système DOIT capturer les mots de passe en utilisant des techniques de masquage, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.10	Capturer les codes porteurs en les masquant	<u>SI les codes porteurs sont utilisés ALORS le système DOIT capturer le code porteur de la CPS en utilisant des techniques de masquage, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
SEC.2.1.11	Gérer la réinitialisation des mots de passe	<u>SI les mots de passe sont utilisés ALORS le système DOIT gérer la réinitialisation de mot de passe en tant que fonction d'administration</u>
SEC.2.1.12	Mettre à jour le mot de passe	<u>SI un mot de passe est initialisé ou réinitialisé ALORS le système DOIT obliger l'utilisateur à mettre à jour son mot de passe à la prochaine ouverture de session</u>
SEC.2.1.13	Présenter une information contextuelle limitée à l'utilisateur durant la phase d'authentification	<u>Le système DOIT présenter une information contextuelle limitée à l'utilisateur durant la phase d'authentification</u>
SEC.2.1.14	Nom d'utilisateur insensible à la casse de caractère	<u>Le système DOIT permettre de saisir des identifiants d'utilisateurs contenant des caractères alphanumériques affichables, sans influence de la casse</u>
SEC.2.1.15	Mot de passe sensible à la casse de caractère	<u>SI les mots de passe sont utilisés ALORS le système DOIT permettre de saisir des mots de passe sensibles à la casse contenant des caractères alphanumériques affichables</u>
SEC.2.1.16	Authentifier les entités autres que des personnes physiques accédant au système	<u>Le système DEVRAIT authentifier les entités autres que des personnes physiques accédant au système, conformément aux pratiques du secteur, à la politique de l'organisation, à la réglementation</u>
Traçabilité des connexions et des actions		
	Non répudiation	Limite la capacité d'un utilisateur du système à répudier la production, l'émission ou la réception de donnée par cet utilisateur
	Protection de l'accès aux traces	Traçabilité des enregistrements clés, de la sécurité, des événements systèmes et des événements cliniques
	Gestion des traces des événements utilisateurs	Gestion des traces utilisateur en capturant qui, quoi, où, quand
	Gestion des traces des événements système	Gestion des traces système en capturant qui, quoi, où, quand
SEC.3.4.3	Tracer les opérations de sauvegarde et restauration	<u>Le système DEVRAIT permettre de tracer les opérations de sauvegarde et de restauration</u>
SEC.3.4.4	Tracer l'accès au système et son utilisation	<u>Si le logiciel est commercialisé en mode SaaS ALORS le système DOIT permettre de tracer les accès au système et à ses données ainsi que les opérations effectuées</u>
SEC.3.4.5	Tracer les accès en télémaintenance	<u>Le système DOIT permettre de tracer les accès en télémaintenance et les opérations effectuées lors de ces accès</u>

6.4. E-mail de sollicitation envoyé via les URPS

Madame, monsieur,

Je réalise un travail de thèse de médecine générale, sur le thème de l'informatique au cabinet médical.

Je suis conscient que ce sujet n'est pas le cœur de notre métier, mais nous utilisons l'outil informatique au quotidien au cabinet, et nos données médicales sont sensibles et vulnérables.

Plus précisément, mon sujet s'intitule :

<< **Protection** contre les **pertes de données** en **médecine générale** :
comment les **données informatisées médicales** sont-elles **protégées**,
sauvegardées et **archivées** par les **médecins généralistes français** ? >>

Je sollicite votre aide pour répondre au questionnaire en ligne me permettant d'appréhender la réalité du terrain.

Pas besoin d'avoir de grandes notions d'informatique !

Ce questionnaire est :

- **Accessible à tous** : explications concises à chaque question
- **Court** : testé à **6 min** en moyenne, entre 3 et 8 minutes selon les réponses
- **Anonyme**
- **Conditionnel** : s'adapte au fur et à mesure de vos réponses → une réponse "je ne sais pas" raccourcit la durée totale du questionnaire, et me donne tout de même de précieuses informations.

Il est disponible à ces deux adresses :

- <https://maguypaul.ovh/limesurvey/index.php/10718>

ou

- <https://micloud.ovh/limesurvey/index.php/10718>

Je vous remercie d'avance pour votre participation.

MAGUY Paul,

médecin généraliste remplaçant (région Auvergne-Rhône-Alpes, faculté Lyon Est)

limesurvey.maguy@laposte.net

6.5. Récépissé de déclaration simplifiée CNIL



3 Place de Fontenoy - 75334 PARIS Cedex 07
T. 01 53 73 22 22 - F. 01 53 73 22 00
www.cnil.fr

RÉCÉPISSÉ

Monsieur MAGUY Paul
14C RUE DU DOCTEUR REBATEL
69003 LYON

DÉCLARATION NORMALE

Numéro de déclaration

2181547 v 0

du 15 mai 2018

A LIRE IMPÉRATIVEMENT

La délivrance de ce récépissé atteste que vous avez transmis à la CNIL un dossier de déclaration formellement complet. Vous pouvez désormais mettre en oeuvre votre traitement de données à caractère personnel.

La CNIL peut à tout moment vérifier, par courrier, par la voie d'un contrôle sur place ou en ligne, que ce traitement respecte l'ensemble des dispositions de la loi du 6 janvier 1978 modifiée en 2004. Afin d'être conforme à la loi, vous êtes tenu de respecter tout au long de votre traitement les obligations prévues et notamment :

- 1) La définition et le respect de la finalité du traitement,
- 2) La pertinence des données traitées,
- 3) La conservation pendant une durée limitée des données,
- 4) La sécurité et la confidentialité des données,
- 5) Le respect des droits des intéressés : information sur leur droit d'accès, de rectification et d'opposition.

Pour plus de détails sur les obligations prévues par la loi « informatique et libertés », consultez le site internet de la CNIL : www.cnil.fr

Organisme déclarant

Nom : Monsieur MAGUY Paul

Service :

Adresse : 14C RUE DU DOCTEUR REBATEL

Code postal : 69003

Ville : LYON

N° SIREN ou SIRET :

Code NAF ou APE :

Tél. : 0615746117

Fax. :

Traitement déclaré

Finalité : QUESTIONNAIRE LINESURVEY AUTO-HÉBERGÉ À MON DOMICILE SUR SERVEUR PERSONNEL, PROPOSÉ AUX MÉDECINS GÉNÉRALISTES LIBÉRAUX INSTALLÉS FRANÇAIS POUR TRAVAIL DE THÈSE SUR LE THÈME : PROTECTION CONTRE LES PERTES DE DONNÉES EN MÉDECINE GÉNÉRALE : COMMENT LES DONNÉES INFORMATISÉES MÉDICALES SONT-ELLES PROTÉGÉES, SAUVEGARDÉES ET ARCHIVÉES PAR LES MÉDECINS GÉNÉRALISTES FRANÇAIS ? ETUDE STATISTIQUE DES RÉSULTATS POUR LA THÈSE. QUESTIONNAIRE ANONYMISÉ.

Fait à Paris, le 15 mai 2018
Par délégation de la commission

Isabelle FALQUE PIERROTIN
Présidente

6.6. Questionnaire imprimable

Protection, sauvegarde et archivage des données informatisées médicales par les médecins généralistes français

Ce questionnaire est destiné aux **médecins généralistes installés** uniquement.

Il s'adapte en longueur en fonction de vos réponses.

Le **temps nécessaire** pour y répondre peut varier **entre 3 et 8 minutes**.

L'utilisation de l'outil informatique pour stocker les informations médicales est aujourd'hui une nécessité pour tous les nouveaux médecins.

La bonne gestion des données médicales n'est possible que si le médecin connaît les enjeux de leur sécurité : les données médicales sont sensibles, fragiles, et doivent être conservées de façon pérenne et sécurisée.

Ce travail vise à explorer les **pratiques des médecins généralistes installés** concernant la **qualité des sauvegardes et archives de leurs données informatisées**.

- MAGUY Paul, médecin remplaçant

Il y a 45 questions dans ce questionnaire

À propos de vous

[]

Êtes-vous une femme ou un homme ? *

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Féminin
☐ Masculin

[] Dans quelle tranche d'âge êtes-vous ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Moins de 31 ans
☐ Entre 31 et 40 ans
☐ Entre 41 et 50 ans
☐ 51 ans et plus

[] De combien d'années date votre installation actuelle ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

`{age.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/260/qid/2364) != "age1"} and ( ! is_empty(age (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/260/qid/2364)))`

Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Moins de 5 ans
☐ Entre 5 et 10 ans
☐ Entre 11 et 20 ans
☐ Plus de 21 ans

[] Exercez-vous seul ou en groupe ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Seul
☐ En groupe

[] Dans quel type de territoire exercez-vous ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Urbain

☐ Rural

Unité urbaine = une commune (ou un ensemble de communes) :

- présentant une zone de bâti continu (pas de coupure de plus de 200 mètres entre deux constructions)
- comptant au moins 2 000 habitants.

Commune rurale = une commune n'appartenant pas à une unité urbaine.

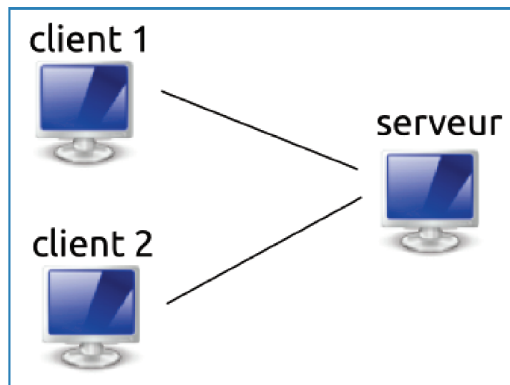
Votre installation informatique

[] Quel est votre type d'installation ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Monoposte
☐ Client/serveur



Monoposte = votre **poste de travail*** est le lieu de stockage de vos données de travail, fonctionne de façon indépendante, et ne joue pas le rôle de serveur pour un autre poste.

Client/serveur = votre poste de travail travaille en réseau avec d'autres postes. Soit il nécessite qu'un autre ordinateur -le serveur- soit allumé pour accéder aux données de travail, soit il est lui-même serveur.

***Poste de travail** : ordinateur situé dans votre cabinet en tant qu'outil de saisie des observations et données médicales informatisées.

[] Concernant votre **progiciel*** : est-il partagé avec au moins un des autres membres du groupement ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :
La réponse était 'En groupe' à la question '4 [solo]' (Exercez-vous seul ou en groupe ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Progiciel : votre logiciel professionnel médical (ex : AxiSanté, HelloDoc, Weda...), ou logiciel de bureautique utilisé à des fins médicales (ex : Microsoft Word, Excel... ou TextEdit, Numbers...)

[] Concernant votre progiciel : s'agit-il d'une version en **cloud*** ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Fonctionnement progiciel en cloud : l'accès à vos données de travail se fait à distance par Internet

[] L'intégralité de vos **données de travail informatisées*** est-elle stockée dans votre progiciel ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Données de travail informatisées : toutes vos données médicales informatisées, que vous complétez au quotidien de votre exercice.

Sont comprises les informations administratives patient, observations médicales, courriers scannés, biologies, photos de suivi dermatologique, etc.

-> Choisissez "Non" si au moins une partie de vos données médicales informatisées est stockée numériquement, mais en dehors de votre progiciel.

[]
{if(cloud.NAOK == "cldo", 'Hormis les éventuelles sauvegardes de votre progiciel au sein de son cloud, existe-t-il une autre', 'Existe-t-il au moins une') } **sauvegarde*** de vos données de santé informatisées ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Sauvegarde : copie intégrale des données de travail, réalisées à un instant T, stockées sur un support différent de celui des données de travail.

Connaissances succinctes

[] Selon vous, le médecin généraliste est-il responsable de la sécurité de ses données informatisées : *

Choisissez la réponse appropriée pour chaque élément :

	Non	Oui	Je ne sais pas
s'il ne fait appel à aucun prestataire ?	☐	☐	☐
s'il en confie la gestion à un prestataire ?	☐	☐	☐

[] Savez-vous dans quel ordinateur sont stockées vos données de travail ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Client/serveur' à la question '6 [reseau]' (Quel est votre type d'installation ?) et La réponse n'était pas 'Oui' à la question '8 [cloud]' (Concernant votre progiciel : s'agit-il d'une version en cloud* ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

[] Connaissez-vous le **chemin d'accès informatique*** de vos données de travail ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse n'était pas 'Oui' à la question '8 [cloud]' (Concernant votre progiciel : s'agit-il d'une version en cloud* ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

Chemin d'accès informatique : position d'un fichier dans un répertoire informatique virtuel.

Exemples en local :

- Windows : "C:\Utilisateur\Mes documents\Données logicielles"

- Apple : "Macintosh HD/Utilisateurs/[Nom d'utilisateur]/Documents"

{if(reseau.NAOK == "rescs", 'Exemple sur serveur : "\SERVEUR\Sauvegardes"', '')}

[] Les données médicales hors progiciel sont-elles incluses dans vos sauvegardes ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse n'était pas 'Oui' à la question '9 [progicielseul]' (L'intégralité de vos données de travail informatisées* est-elle stockée dans votre progiciel ?) et La réponse était 'Oui' à la question '10 [sauvegarde]' ({if(cloud.NAOK == "cldo", 'Hormis les éventuelles sauvegardes de votre progiciel au sein de son cloud, existe-t-il une autre', 'Existe-t-il au moins une')}) sauvegarde* de vos données de santé informatisées ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Le risque informatique et vous

[] Quel niveau de confiance accordez-vous à : *

Choisissez la réponse appropriée pour chaque élément :

	1	2	3	4	5
la sécurité de vos données médicales informatisées	☐	☐	☐	☐	☐

1 = Faible -> 5 = Fort

[] Quel niveau de compétence personnelle pensez-vous avoir concernant : *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '10 [sauvegarde]' ({if(cloud.NAOK == "cldo", 'Hormis les éventuelles sauvegardes de votre progiciel au sein de son cloud, existe-t-il une autre', 'Existe-t-il au moins une')}) sauvegarde* de vos données de santé informatisées ?)

Choisissez la réponse appropriée pour chaque élément :

	1	2	3	4	5
la réalisation de vos sauvegardes	☐	☐	☐	☐	☐
la restauration de vos sauvegardes	☐	☐	☐	☐	☐

1 = Faible -> 5 = Fort

[]
 {if(age.NAOK == "age1", 'Depuis le début de votre installation', 'Au cours des 5 dernières années')}, avez déjà perdu tout ou une partie de vos données médicales informatisées ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

À propos de votre perte de données

Si vous avez subi plusieurs épisodes de pertes de données, on parlera ici **uniquement** de l'épisode le plus récent.

[]Quels ont été les causes de la perte de données ? *

Cochez la ou les réponses

Veuillez sélectionner au moins une réponse

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Inconnue ou non retrouvée
☐ Panne matérielle
☐ Erreur humaine
☐ Dysfonctionnement logiciel
☐ Catastrophe naturelle (inondation, incendie...)
☐ Virus informatique
☐ Vol d'un support
☐ Perte d'un support

[]Au moment de la perte de données, combien de jours de saisie de donnée ont été perdus ? *

Seul un nombre entier peut être inscrit dans ce champ.

Veuillez écrire votre réponse ici :

[]Avez-vous alors un système de sauvegarde ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

[]Avez-vous pu restaurer les données perdues à partir d'une sauvegarde ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '20 [pertesauegarde]' (Avez-vous alors un système de sauvegarde ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

[]Avez-vous pu restaurer la totalité des données perdues ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '21 [perterestauree]' (Avez-vous pu restaurer les données perdues à partir d'une sauvegarde ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

[]Combien de jours de saisie ont été définitivement perdus ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '21 [perterestauree]' (Avez-vous pu restaurer les données perdues à partir d'une sauvegarde ?) et La réponse était 'Non' à la question '22 [perterestaureetotal]' (Avez-vous pu restaurer la totalité des données perdues ?)

Seul un nombre entier peut être inscrit dans ce champ.

Veuillez écrire votre réponse ici :

[] Combien de temps pour retrouver un fonctionnement de travail normal ? (en jours)
*

Seul un nombre entier peut être inscrit dans ce champ.

Veuillez écrire votre réponse ici :

Vos méthodes de sauvegarde

{if(cloud.NAOK == "cldo", 'On parlera ici **uniquement** des sauvegardes **autres que celles du progiciel réalisées au sein de son cloud.**', '')}

[] La réalisation de vos sauvegardes se fait-elle automatiquement ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

[] Qui réalise {if(sauvegardeauto.NAOK == "sao", 'et/ou a automatisé ', '')} vos sauvegardes ? *

Cochez la ou les réponses

Veuillez sélectionner au moins une réponse

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Moi-même
☐ Un autre membre du groupe d'exercice
☐ Un tiers contractuel, hors {if(solo.NAOK == "soloo", 'cabinet', 'groupe d'exercice')} (ex: technicien informatique)
☐ Un tiers autre

[] Votre processus de sauvegarde utilise-t-il l'**outil de sauvegarde intégré*** à votre progiciel ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Outil de sauvegarde intégré : la plupart des progiciels médicaux comportent dans leurs réglages une option pour réaliser facilement une sauvegarde de toutes leurs données.

Vous pouvez cocher « non » si votre progiciel n'a pas cette option, ou si vous ne vous en servez pas.

[] Quelles méthode de sauvegarde utilisez-vous ?
*

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Non' à la question '27 [sauveprogiciel]' (Votre processus de sauvegarde utilise-t-il l'outil de sauvegarde intégré* à votre progiciel ?)

Cochez la ou les réponses

Veuillez sélectionner au moins une réponse

Veuillez choisir toutes les réponses qui conviennent :

- ☐ 1) Un utilitaire de sauvegarde intégré au système d'exploitation
☐ 2) Un programme tiers
☐ 3) Une copie manuelle des données du progiciel
☐ Je ne sais pas

1) Windows : Sauvegarder et restaurer Windows 7, Sauvegarde windows 10
Mac : TimeMachine

2) Windows : Acronis, AOMEI, EaseUS Todo Backup...
Mac : CarbonCopy, SuperDuper!, Chronosync...

3) en allant chercher dans l'explorateur de fichier les données à sauvegarder

[]

Envoyez-vous {if(cloud.NAOK == "cldo",'une sauvegarde supplémentaire sur un **cloud tiers*** ','une sauvegarde sur un **cloud tiers*** ')} *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non
☐ Je ne sais pas

Cloud tiers : Services de stockage en cloud sans rapport avec votre progiciel. Exemples :

- Dropbox
- iCloud (Apple)
- Amazon Drive
- Hubic
- Google Drive
- OneDrive (Microsoft)
- pCloud
- Certain Safe

[]

{if(!((sauvegardecloud.NAOK == "sco")or(cloud.NAOK == "cldo")), 'P','Hormis ')}
{if(cloud.NAOK == "cldo", 'les sauvegardes de votre progiciel en cloud, ','')}
{if(((sauvegardecloud.NAOK == "sco")and(cloud.NAOK == "cldo")), 'et ','')}
{if(sauvegardecloud.NAOK == "sco", 'les sauvegardes réalisées sur un cloud tiers, ','')}
{if(!((sauvegardecloud.NAOK == "sco")or(cloud.NAOK == "cldo")), ', 'p')}récisez
le nombre de sauvegardes par type de **support physique***. *

Merci d'indiquer au moins un support de sauvegarde.

	Stocké au cabinet médical	Stocké hors cabinet médical
Clé USB		
Optique = CD/DVD gravé		
Disque dur externe		
↳ dont SSD*		
Disque dur en réseau (NAS)		

Support physique (informatique) : dispositif physique utilisé pour stocker les données informatiques.

Si vous avez plus de 6 supports dans une catégorie, merci d'entrer "6".

{if(((sauvegardesupports_sddext_slocal)>0)or((sauvegardesupports_sddext_sext)>0),"Si vous ne savez pas ce qu'est un **SSD**, laisser à 0. ','")}

[]

{sauvegardesupports_susb_slocal=if(is_empty(sauvegardesupports_susb_slocal.NAOK),'0',sauvegardesupports_susb_slocal.NAOK)}
{sauvegardesupports_sopt_slocal=if(is_empty(sauvegardesupports_sopt_slocal.NAOK),'0',sauvegardesupports_sopt_slocal.NAOK)}
{sauvegardesupports_sddext_slocal=if(is_empty(sauvegardesupports_sddext_slocal.NAOK),'0',sauvegardesupports_sddext_slocal.NAOK)}
{sauvegardesupports_sssd_slocal=if(is_empty(sauvegardesupports_sssd_slocal.NAOK),'0',sauvegardesupports_sssd_slocal.NAOK)}
{sauvegardesupports_snas_slocal=if(is_empty(sauvegardesupports_snas_slocal.NAOK),'0',sauvegardesupports_snas_slocal.NAOK)}
{sauvegardesupports_susb_sext=if(is_empty(sauvegardesupports_susb_sext.NAOK),'0',sauvegardesupports_susb_sext.NAOK)}
{sauvegardesupports_sopt_sext=if(is_empty(sauvegardesupports_sopt_sext.NAOK),'0',sauvegardesupports_sopt_sext.NAOK)}
{sauvegardesupports_sddext_sext=if(is_empty(sauvegardesupports_sddext_sext.NAOK),'0',sauvegardesupports_sddext_sext.NAOK)}
{sauvegardesupports_sssd_sext=if(is_empty(sauvegardesupports_sssd_sext.NAOK),'0',sauvegardesupports_sssd_sext.NAOK)}
{sauvegardesupports_snas_sext=if(is_empty(sauvegardesupports_snas_sext.NAOK),'0',sauvegardesupports_snas_sext.NAOK)}

[] Vos sauvegardes sur support physique sont-elles **chiffrées*** ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

((sum(sauvegardesupports_susb_slocal.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_susb_sext.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sopt_slocal.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sopt_sext.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sddext_slocal.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sddext_sext.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sssd_slocal.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_sssd_sext.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_snas_slocal.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385),sauvegardesupports_snas_sext.NAOK (/limesurvey/index.php/admin/questions/sa/view/surveyid/10718/gid/263/qid/2385))) != 0)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui, toutes
- ☐ Oui, une partie
- ☐ Non
- ☐ Je ne sais pas

Chiffrement : procédé de cryptographie rendant impossible la lecture d'un document (ou support physique) sans une clé de déchiffrement.

[]

Votre sauvegarde sur cloud tiers est-elle chiffrée ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '29 [sauvegardecloud]' (Envoyez-vous {if(cloud.NAOK == "cldo","une sauvegarde supplémentaire sur un cloud tiers* ?","une sauvegarde sur un cloud tiers* ?")})

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non
- ☐ Je ne sais pas

[]

Concernant la sauvegarde sur cloud tiers : est-elle chiffrée avant son envoi sur le cloud tiers ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '33 [cloudtierschiffre]' (Votre sauvegarde sur cloud tiers est-elle chiffrée ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non
- ☐ Je ne sais pas

[]

Vérifiez-vous la réussite du processus de sauvegarde ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse n'était pas 'Oui' à la question '8 [cloud]' (Concernant votre progiciel : s'agit-il d'une version en cloud* ?)

Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Oui, et pour cela ...
- ☐ - je consulte le journal de sauvegarde
- ☐ - je vérifie l'existence des fichiers sauvegardés
- ☐ Non

Vos méthodes : restauration de sauvegarde

[]

Avez-vous déjà tenté de **restaurer*** une sauvegarde, en dehors du contexte de perte de données ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non

Restauration : utilisation d'une sauvegarde pour restaurer les données de travail.

[]

A quelle fréquence tentez-vous cette restauration ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '36 [restaurationtestee]' (Avez-vous déjà tenté de restaurer* une sauvegarde, en dehors du contexte de perte de données ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Une seule fois en tout
- ☐ Moins d'une fois par an
- ☐ Une fois par an ou plus souvent

[]

Sur quelle ordinateur le (ou les) essai(s) de restauration est-il fait ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '36 [restaurationtestee]' (Avez-vous déjà tenté de restaurer* une sauvegarde, en dehors du contexte de perte de données ?)

Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ ordinateur séparé du réseau du cabinet (hors-système)
- ☐ {if(reseau.NAOK == "rescs", 'directement sur mon poste de travail', 'directement sur le serveur')}

Vos méthodes : en cas de panne

[] Connaissez-vous la marche à suivre détaillée pour : *

Choisissez la réponse appropriée pour chaque élément :

	Oui	Non
continuer à travailler malgré la panne (plan de continuité)	☐	☐
restaurer les données à partir d'une sauvegarde (plan de restauration)	☐	☐

Vos archives informatiques

ATTENTION : "**sauvegarde**" ≠ "**archive**" !

Pour simplifier : une **archive** est un exemplaire de sauvegarde auquel on ne touche plus, stocké à l'abri, pour conservation à long terme.

[] Réalisez-vous une **archive** de vos dossiers médicaux informatisés ? *

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
- ☐ Non

[] Pendant combien d'années conservez-vous un dossier médical informatique archivé ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '40 [archives]' (Réalisez-vous une archive de vos dossiers médicaux informatisés ?)

Seul un nombre entier peut être inscrit dans ce champ.

Veuillez écrire votre réponse ici :

ans

[] Sur quel type de support informatique sont réalisées vos archives ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '40 [archives]' (Réalisez-vous une archive de vos dossiers médicaux informatisés ?)

Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Clé USB
- ☐ Disque dur externe
- ☐ ☐ , dont au moins 1 **SSD***
- ☐ CD/DVD/Blu-Ray standard
- ☐ DVD/Blu-Ray **M-Disc***
- ☐ Bande magnétique
- ☐ Sur internet (Cloud)

M-Disc : Disque enregistrable optique spécifique à longue durée de vie.

{ff((archivessupport_asxhd.NAOK == "Y"), "Si vous ne savez pas ce qu'est un **SSD**, ne pas cocher cette case.", "")}

[] Où se situent physiquement les supports de vos archives informatisées ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '40 [archives]' (Réalisez-vous une archive de vos dossiers médicaux informatisés ?)

Cochez la ou les réponses

Veuillez choisir toutes les réponses qui conviennent :

- ☐ Au cabinet
- ☐ Hors cabinet

[] Recopiez-vous vos archives pour en préserver la durabilité ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '40 [archives]' (Réalisez-vous une archive de vos dossiers médicaux informatisés ?)

Veuillez sélectionner une réponse ci-dessous

Veuillez sélectionner une seule des propositions suivantes :

- ☐ Oui
☐ Non

[] A quelle fréquence réalisez-vous la recopie des archives ? *

Répondre à cette question seulement si les conditions suivantes sont réunies :

La réponse était 'Oui' à la question '44 [archivesrecopie]' (Recopiez-vous vos archives pour en préserver la durabilité ?)

Seul un nombre entier peut être inscrit dans ce champ.

Veuillez écrire votre réponse ici :

Tous les

ans

Un grand **merci** à vous !

Si vous souhaitez être tenu au courant des résultats, ou pour toute autre remarque, vous pouvez m'écrire à :

limesurvey.maguy@laposte.net (mailto:limesurvey.maguy@laposte.net?
subject=Questionnaire%20%22Protection%2C%20sauvegarde%20et%20archivage%20des%20donn%C3%A9es%20in

Rappel : Ce questionnaire Limesurvey est **anonyme**.

L'enregistrement de vos réponses à ce questionnaire ne contient aucune information permettant de vous identifier, à moins que l'une des questions ne vous le demande explicitement.

Envoyer votre questionnaire.

Merci d'avoir complété ce questionnaire.

6.7. Définitions de notions informatiques

6.7.1. Définitions présentes dans le questionnaire

Les définitions suivantes sont présentées suivant leur ordre d'apparition au fil du questionnaire. Elles sont volontairement courtes et simplifiées, pour permettre une accessibilité des termes techniques sans entrer trop dans le détail.

- **Monoposte** = votre poste de travail* est le lieu de stockage de vos données de travail, fonctionne de façon indépendante, et ne joue pas le rôle de serveur pour un autre poste.
- **Client/serveur** = votre poste de travail travaille en réseau avec d'autres postes. Soit il nécessite qu'un autre ordinateur -le serveur- soit allumé pour accéder aux données de travail, soit il est lui-même serveur.
- **Poste de travail** : ordinateur situé dans votre cabinet en tant qu'outil de saisie des observations et données médicales informatisées.
- **Progiciel** : votre logiciel professionnel médical (ex : AxiSanté, HelloDoc, Weda...), ou logiciel de bureautique utilisé à des fins médicales (ex : Microsoft Word, Excel... ou TextEdit, Numbers...)
- **Fonctionnement progiciel en cloud** : l'accès à vos données de travail se fait à distance par Internet.
- **Données de travail informatisées** : toutes vos données médicales informatisées, que vous complétez au quotidien de votre exercice. Sont comprises les informations administratives patient, observations médicales, courriers scannés, biologies, photos de suivi dermato, etc.

- **Sauvegarde** : copie intégrale des données de travail, réalisées à un instant T, stockées sur un support différent de celui des données de travail.
- **Chemin d'accès informatique** : position d'un fichier dans un répertoire informatique virtuel.
Exemples en local :
 - Windows : "C:\Utilisateur\Mes documents\Données logicielles"
 - Apple : "Macintosh HD/Utilisateurs/[Nom d'utilisateur]/Documents"
 - Exemple sur serveur : "\\SERVEUR\Sauvegardes\" sous Windows, "afp:///serveur/sauvegardes" sous MacOS.
- **Outil de sauvegarde intégré** : la plupart des progiciels médicaux comportent dans leurs réglages une option pour réaliser facilement une sauvegarde de toutes leurs données.
- **Cloud tiers** : Services de stockage en *cloud* sans rapport avec votre progiciel. Exemples :
 - Dropbox
 - Google Drive
 - iCloud (Apple)
 - OneDrive (Microsoft)
 - Amazon Drive
 - pCloud
 - Hubic
 - Certain Safe
- **Support physique (informatique)** : dispositif physique utilisé pour stocker les données informatiques.
- **Chiffrement** : procédé de cryptographie rendant impossible la lecture d'un document (ou support physique) sans une clé de déchiffrement.
- **Restauration** : utilisation d'une sauvegarde pour restaurer les données de travail.
- **M-Disc** : *Millennial Disc*, en français "Disque Millénaire". Disque enregistrable optique spécifique à longue durée de vie. Existe au standard DVD ou Blu-Ray. Nécessite un graveur compatible, support non réinscriptible. Coût à l'unité du support largement supérieur à celui des supports optiques classiques.

6.7.2. Définitions techniques supplémentaires

- **Authentification à Facteurs Multiples** : contrairement aux méthodes d'authentifications classiques reposant sur un seul facteur, nécessité d'apporter plusieurs facteurs en simultanée permettant de renforcer la sécurité et la fiabilité de l'identité de la personne qui s'authentifie. Les facteurs possibles sont :
 - Quelque chose que l'on connaît : mot de passe, phrase...
 - Quelque chose que l'on a : un objet unique faisant office de clé, détenu par et confié uniquement à la personne cherchant à s'identifier. Exemple : la CPS des médecins, clé de sécurité physique U2F (Universal Second Factor) type Google Titan...
 - Quelque chose que l'on est : par la mesure d'un paramètre biométrique, exemple reconnaissance faciale, lecteur d'empreinte digitale...
- **Déduplication** : Procédé d'analyse des données à un niveau inférieur au fichier, permettant de ne stocker que les éléments qui n'ont pas déjà été stockés sur le média de sauvegarde.

Lors de l'ajout de données à une sauvegarde dédoublée, les nouvelles données sont lues, séparées en blocs, qui sont eux comparés aux blocs déjà stockés. Seuls les nouveaux blocs sont stockés.

Les blocs déjà existants sont créés en tant que pointeurs vers les données déjà stockées. Ceci réduit la quantité de données en ne stockant les blocs qu'une fois pour chaque passage ultérieur de données.

La réduction de quantité de données est mesurée grâce au ratio entre la quantité de données traitée et la quantité de données stockée, appelée ratio de déduplication. Habituellement, un bon ratio de déduplication approche des 10:1, et peut souvent atteindre 15:1. (traduit depuis [\(4\)](#))

- **Docker** : Outil logiciel qui peut emballer une application et ses dépendances dans un conteneur isolé, qui pourra être exécuté sur n'importe quel serveur Linux. Il ne s'agit pas de virtualisation, mais de conteneurisation, une forme plus légère qui s'appuie sur certaines parties de la machine hôte pour son fonctionnement. Cette approche permet d'accroître la flexibilité et la [portabilité d'exécution](#) d'une application, laquelle va pouvoir tourner de façon fiable et prédictible sur une grande variété de machines hôtes. [\(49,50\)](#)
- **Duplicati** : Logiciel de sauvegarde qui planifie et stocke une copie complète ou incrémentielle des fichiers à conserver. La copie de sauvegarde peut être compressée, chiffrée et stockée localement, en ligne, depuis des services de *cloud computing* ou vers des serveurs de fichiers distants. Utilise la [déduplication](#). [\(51\)](#)
- **ECC** : *Error-Correcting Code Memory*, ou mémoire à code correcteur d'erreurs. Un type de mémoire vive qui contient un code correcteur lui permettant de détecter et de corriger les types les plus courants de corruption de données. Ce type de mémoire est utilisé dans les ordinateurs où la corruption de données ne peut être tolérée en aucun cas, comme pour les calculs scientifiques ou financiers. [\(52\)](#)
- **Instantané (*snapshot* en anglais)** : En informatique, état d'un système à un instant donné. Diffère d'une sauvegarde traditionnelle du système par le fait que les données à sauvegarder ne sont copiées sur le périphérique de sauvegarde qu'au moment où elles sont modifiées sur le système dont on crée l'instantané. [\(53\)](#)
- **Journal informatique (*log file* ou *log* en anglais)** : Fichier contenant les enregistrements séquentiels de tous les événements affectant un processus particulier (application, activité d'un réseau informatique...) [\(54\)](#).
- **Machine virtuelle** : Selon Microsoft [\(65\)](#), "une machine virtuelle est un fichier informatique, généralement appelé image, qui se comporte comme un ordinateur réel. En d'autres termes, il s'agit d'un ordinateur créé à l'intérieur d'un ordinateur. Elle s'exécute dans une fenêtre, comme tout autre programme, en offrant à l'utilisateur final une expérience identique à celle qu'il aurait sur le système d'exploitation hôte. La machine virtuelle est placée dans un « bac à sable » qui l'isole du reste du système, de sorte que les logiciels installés sur la machine virtuelle ne peuvent ni s'échapper, ni modifier l'ordinateur hôte."

- **Mémoire Flash** : Type de puce mémoire relativement compact et peu énergivore, présent dans les clés USB et disques SSD. Utilise comme cellule de base un transistor MOS possédant une grille flottante enfouie au milieu de l'oxyde de grille, entre le canal et la grille. L'information est stockée grâce au piégeage d'électrons dans cette grille flottante. Ce type de mémoire diffère donc fondamentalement du disque dur classique, à plateaux métalliques magnétiques en rotation.
- **Open-source** : “code source ouvert” en français. Logiciels dont la licence respecte les possibilités de libre redistribution, d'accès au code source et de création de travaux dérivés. Mis à la disposition du grand public, ce code source est généralement le résultat d'une collaboration entre programmeurs. [\(55\)](#)
- **Pérennité des supports** : durée de vie d'un support de stockage, donnée cruciale dans le choix du support de sauvegarde ou d'archivage. On considère ici à la fois la **pérennité physique du support** (sa capacité à continuer à fonctionner dans le temps) et sa **durée de rétention d'information**. Dépend des conditions de stockage, de l'utilisation active ou en tant qu'archive, et de la nature du support physique. [\(39\)](#)
 - Disque dur classique : environ 10 ans en utilisation non intensive de par l'usure des pièces mobiles. Susceptible aux chocs, à une manipulation en cours de fonctionnement. Également sensible à une démagnétisation, spontanée ou induite par champs magnétiques extérieurs. Dans le cas d'un disque d'archive, laissé éteint sans alimentation électrique, il est considéré comme sécurisé de réécrire les données qu'ils contiennent tous les deux ans au minimum. Cette opération est recommandée tous les ans, afin d'éviter de façon quasi certaine une perte de données, en re magnétisant les secteurs contenant les données.
 - [Mémoire Flash](#) - clés USB et SSD : varie en fonction de l'ancienneté du procédé de fabrication. Dépend surtout de deux facteurs :
 - Le nombre de cycles d'écriture effectué sur les cellules de stockage de la mémoire Flash. La plupart des SSD modernes emploient une stratégie d'*over-provisioning*: la capacité totale accessible à l'utilisateur (exemple 480Go) est inférieure à la quantité de données effectivement stockable sur le disque (exemple 512Go). En cas d'erreur d'écriture sur une cellule, une puce de gestion d'usure à l'écriture met en service une cellule initialement inaccessible pour remplacer celle hors-service. Ce système permet initialement de pallier totalement la perte de secteurs. Une fois les cellules excédentaires toutes recrutées et si l'usure continue, une diminution de la capacité totale apparente du disque apparaît, devant indiquer à l'utilisateur qu'il est temps de remplacer le SSD. Cette technologie n'est souvent pas présente sur les clés USB qui ont un contrôleur mémoire rudimentaire sans gestion de répartition d'usure.
 - La fuite d'électrons présents dans une cellule chargée électriquement. Plus la température de stockage de la mémoire Flash est élevée, plus ce phénomène se produit fréquemment, finissant par décharger certaines cellules de façon trop importante, transformant un 1 et un 0. Ce phénomène induit une corruption de

données, sans détruire la mémoire flash elle-même. Selon une présentation de 2015 par le président de JEDEC (entreprise créant et gérant les standards de fabrication des SSDs) : pour un SSD en fin de vie (aux cellules déjà usées), stocké en dessous des 30°C, la corruption d'information met entre 1 à 2 ans à se produire [\(17\)](#).

- Supports optiques : susceptibles à la dégradation de leur couche métallique ou organique de support de données, pouvant se produire par exposition aux UVs, aux changements de température, au taux d'humidité... et la qualité du support initial. Estimée entre 5 à 10 ans si conservation dans de bonnes conditions : à l'abri de la lumière, ~20°, pas d'humidité.
- [M-Disc](#) : serait plus robuste que les supports optiques classiques. Des tests effectués en conditions extrêmes montrent selon les marques au pire une longévité identique à des supports classiques, au mieux supérieure (47). Leur promotion en est faite sous l'affirmation que le disque retiendra les données sur le temps d'une vie.
- **Rotation des supports** : sauvegarde sur plusieurs supports physiques, avec une alternance dans le temps du support employé, afin de conserver une sauvegarde intégrale correspondant à plusieurs dates sur plusieurs supports. Cette façon de procéder a été rendue obsolète par l'avènement des [sauvegardes incrémentales](#).
- **Redondance** : stockage de la même information en plusieurs exemplaires, chacun stocké sur un support physique distinct. Ceci permet de réduire le risque de perte de données, en réduisant l'impact de la perte d'un support physique.
 - **RAID** : *Redundant Array of Independent Disks* [\(56\)](#) : en français, "Regroupement Redondant de Disques Indépendants". Plusieurs modes d'association de disques durs existent, désignés par un numéro. Le RAID 0 est à part car non redondant. Ce mode est "également connu sous le nom d'«entrelacement de disques» [...], permettant d'augmenter significativement les performances de la grappe en faisant travailler plusieurs disques durs en parallèle." Permet, selon le mode utilisé et si supérieur au RAID 0, d'avoir une répartition redondante intelligente sur plusieurs disques durs des fichiers à sauvegarder. Exemple du RAID 6 (comportant au minimum 4 disques) : tolérance de perte de un ou deux disques durs sans perte de données. Ensuite, une perte de plus de deux disques conduit à la perte de l'ensemble des données sauvegardées.
- **Règle des 3-2-1** : Règle de bon sens régissant la base d'une stratégie de sauvegarde résiliente, en trois parties :
 - Conserver au moins 3 copies de vos fichiers : le fichier principal (ici, les données de travail) et deux sauvegardes. Protège contre un risque de corruption simultanée des données de travail et d'une des sauvegardes. Dans ce cas, il reste au moins un exemplaire des données utilisable.
 - Stocker les fichiers sur 2 types de supports différents : par exemple, un disque dur associé à un support optique. Protège contre les risques liés à un défaut de pérennité du support (défaut physique ou durée de rétention dépassée)

- Conserver 1 copie de sauvegarde hors-site : justifié par le proverbe “on ne met pas tous ses œufs dans le même panier”. En cas d'accident local comme un incendie, disposer d'une sauvegarde hors-site évite la perte définitive de données.

Suivre ces trois directives permet d'assurer une résilience relativement importante du système de sauvegarde. [\(57,58\)](#)

Dans le cas d'un progiciel hébergé en *cloud*, il faudrait même respecter une **configuration en 3-2-2** [\(59\)](#). Dans ce cas, il y aurait au moins deux sauvegardes en *cloud* situées dans deux sites géographiquement distincts. Ce type de sauvegarde est allégué par plusieurs éditeurs de progiciel *cloud*. Cependant, notre travail se concentrait sur les pratiques des généralistes, et pas sur la vérification des pratiques des hébergeurs de progiciel en *cloud*.

- **Sauvegarde incrémentale** : Mode de sauvegarde permettant d'avoir, sur le même support physique, plusieurs exemplaires de sauvegarde à des dates différentes. Permet :
 - une conservation des éléments supprimés, rendant récupérable un fichier effacé par mégarde théoriquement jusqu'à la date de la plus ancienne sauvegarde
 - une sauvegarde plus rapide lorsque les quantités de données rajoutées chaque jour sont peu importantes (ajout à la sauvegarde uniquement des éléments nouveaux)
 - une sauvegarde de taille (en quantité de données totale) plus réduite que plusieurs sauvegardes intégrales sur plusieurs supports
 - une réduction des coûts et du temps utilisé par rapport à un système de rotation des supports
- **SoC : System on a Chip** (Système sur une Puce). “Système informatique complet embarqué sur une puce, pouvant comprendre de la mémoire, un ou plusieurs microprocesseurs, des périphériques d'interface, ou tout autre composant nécessaire à la réalisation de la fonction attendue.” [\(60\)](#)
- **SSD : Solid State Drive**. Un type de disque dur apparu en 2006. Ils contiennent le même type de mémoire que celle d'une clé USB, de type 'flash'. De par leur relative nouveauté, ils commencent à être intégrés dans les ordinateurs domestiques depuis seulement quelques années. Ils présentent certains avantages par rapport aux disques classiques : l'absence de pièce mobile les rend moins énergivores et plus solides. Ainsi, le risque de perte de donnée lors de vibrations, chocs ou usure des pièces motrices est quasi nul. Par rapport à un disque dur classique, ils sont plus légers car sans volumineuse pièce de métal, et plus performants : leur débit est souvent plus de 10 fois supérieur en lecture, encore davantage en écriture. La latence d'accès aux données est quasi nulle car il n'y a pas besoin d'aligner une tête de lecture sur un plateau. Ils ne sont pas exempts de défauts : le prix au gigaoctet reste encore à l'avantage du disque mécanique, mais cet écart tend à se réduire avec la maturation de la technologie. Ils restent sujets à la même forme d'usure que les clés USB : chaque écriture sur leurs puces mémoires 'use' les cellules de stockage (cf. [Pérennité des supports](#)). Cependant, la durée de vie d'un disque SSD moderne, en usage quotidien bureautique, devrait dépasser celle d'un disque dur classique par le moindre risque d'usure

mécanique. La technologie est encore récente et laisse sceptique sur la durée de vie totale prévue d'un disque SSD. Il est donc d'usage commun de les utiliser pour stocker les données du système d'exploitation, et de conserver comme moyen de stockage de masse durable un disque dur classique.

- **Thin client** : “Client léger” en français. Peut désigner du matériel ou un logiciel. Au sens matériel, un client léger est un ordinateur qui, dans une architecture client-serveur, n'a presque pas de logique d'application. Il dépend donc surtout du serveur central pour le traitement. Un client léger peut se contenter d'une machine minimaliste en termes de matériel. La logique qui prévaut au déploiement de clients légers est une logique essentiellement économique. Il s'agit de réduire le coût total de possession et de gestion. Il ne faut toutefois pas oublier une certaine dimension écologique dans le cas du recyclage de vieux ordinateurs en clients légers matériels. [...] Dans le cadre des clients matériels, les clients légers sont beaucoup moins sujets aux pannes comparativement aux PC : ils fonctionnent avec un processeur qui chauffe peu et n'ont généralement pas de disque dur. De plus, en cas de nécessité d'évolution, il suffit de changer le serveur central. D'autre part, l'administration est beaucoup plus centralisée, et nécessite donc moins de personnel, mais plus qualifié [\(61\)](#).
- **Verbatim** : Selon le dictionnaire Larousse, “*Reproduction intégrale des propos prononcés par l'interviewé ; compte rendu fidèle*” [\(62\)](#). Dans le cas de données informatiques, conservation parfaitement à l'identique des données sources, sans erreur ou corruption de données après copie.

6.8. Choix du support logiciel du questionnaire

Pour la mise en ligne du questionnaire, plusieurs solutions se sont présentées au début de notre travail. Google Forms, bien que pratique et performant, présentait - comme évoqué précédemment - le désavantage d'une confidentialité nulle. Ce sentiment a par la suite été confirmé par la faculté de médecine de Lyon qui a interdit en 2018 à ses thésards d'utiliser Google Forms, pour ces raisons précises.

Nous nous sommes alors tournés vers les alternatives [open-source](#), en particulier le logiciel Limesurvey. Il s'agit d'une plateforme de questionnaire open-source. Il est possible avec ce logiciel d'employer leur service, appelé “Limeservice” puis “Limesurvey Professional”, qui est un hébergement payant sur leurs serveurs.

Cet hébergement permet de se dégager d'un risque d'exploitation des données recueillies par Google, mais a un coût : le pack Basic coûtait 29€/mois au moment de la création du questionnaire.

Il est possible de réaliser une installation de Limesurvey sur n'importe quelle machine tournant sous Linux ou Windows, appelée Limesurvey CE (Community Edition), en la téléchargeant depuis leur site Internet [\(63\)](#).

Limesurvey nécessite une base de données pour fonctionner, et requiert donc une installation d'une application de base de données type MySQL, PostgreSQL ou Microsoft SQL (sous Windows).

Une fois l'installation réalisée, Limesurvey est accessible via une URL locale. On a alors affaire à une interface web permettant de créer et éditer un questionnaire, qui est par la suite rendu accessible en ligne

6.9. Hébergement physique du questionnaire

Il est théoriquement également possible de louer un service chez un fournisseur comme OVH, mais c'est une solution payante au mois, d'autant plus chère que le serveur hébergeur fourni est puissant. La puissance de calcul et la quantité de stockage requise pour héberger une installation Limesurvey étant relativement faibles, nous avons initialement sollicité le département informatique de la faculté. Malheureusement, celui-ci n'a pas pu offrir cette possibilité d'hébergement. Nous avons donc été contraints de faire le choix d'un auto-hébergement.

Afin de renforcer la fiabilité du questionnaire, il nous a semblé judicieux d'avoir une installation redondante, sur deux systèmes indépendants. Ainsi, en cas de défaillance de l'une des deux installations, l'autre reste accessible via son adresse internet spécifique.

La charge informatique supposée était d'au grand maximum une centaine de participants en simultané. Les capacités d'un seul RPi auraient été en soi théoriquement suffisantes. Cependant, cette installation Limesurvey auto-hébergée a été une première pour le thésard. Il aurait été imprudent de faire reposer intégralement la capacité de collecte des réponses sur une seule machine, d'où un besoin d'une installation **redondante**.

Concernant la visibilité sur Internet, nous avons opté pour l'offre OVH afin d'acquérir le nom de domaine "maguypaul.ovh". Cette extension en ".ovh" a été choisie pour des raisons de coût réduit : 0,99€/an, par rapport aux extensions plus populaires en ".fr" (5,49€/an) ou en ".com" (7,99€/an).

Une redirection est donc établie entre l'adresse IP de la BBox du thésard et cette adresse internet, et diffusée sur les serveurs DNS d'OVH. En somme, ce procédé permet donc de rendre l'adresse IP de la BBox via l'adresse internet "maguypaul.ovh", et donc le questionnaire à l'adresse "maguypaul.ovh/limesurvey", accessibles de partout sur la planète, depuis n'importe quel terminal relié à Internet. On obtient au final un accès simple, confidentiel, sécurisé et accessible en permanence des répondants au questionnaire.

L'installation redondante a été faite sur un serveur personnel physiquement situé dans la ville d'Antibes, appartenant à M. MICHEL Jérôme, accessible via l'URL "micloud.ovh/limesurvey".

6.10. Hébergement logiciel : installations Limesurvey

Afin de comparer et comprendre la circulation des données dédiées à Limesurvey sur chacune des deux installations, leurs schémas emploient une légende similaire.

L'installation "Micloud", au domicile de M. MICHEL Jérôme, située à Antibes, utilise un compartimentage à l'aide de [machines virtuelles](#). L'installation "Rebatel" au domicile du thésard n'en

utilise pas : la virtualisation sur le processeur ARM du RPi n'est pas possible. D'où un emploi plus large de [Docker](#), permettant de compartimenter facilement les programmes.

Concernant Limesurvey : il n'y a pas eu d'application des mises à jour aux versions les plus récentes sorties au fur et à mesure du recueil de données afin de limiter les risques de perte de fonctionnement si des composants devenaient incompatibles entre eux. Par contre, des mises à jour manuelles de Raspbian via une commande `sudo apt update && sudo apt upgrade -y -qq` ont été effectuées, en s'assurant de l'absence de message d'erreur.

6.11. Installation principale "Rebatel"

L'installation au domicile du thésard (Rue du Docteur Rebatel, Lyon) ayant permis l'hébergement du questionnaire limesurvey comprenait :

- Une BBox (box internet Bouygues, fibrée 1Gbit/s)
- Deux RPi modèle 3B
 - Raspi1 :
 - Proxy Nginx d'accueil des flux sur le port 443
 - Limesurvey sur serveur Apache2
 - Base de données MySQL
 - Un Docker LetsEncrypt (et son proxy Nginx dédié) redirigeant les demandes sur port 80 sur un flux chiffré sur port 443
 - Un Docker Duplicati pour la sauvegarde de la base de données
 - Raspi2 :
 - Base de données MySQL dupliquée
 - Docker OpenVPN pour accès à distance sécurisé au réseau local et interfaces d'administration
 - Autres Dockers sans rapport avec le travail de thèse
- Un NAS WD MyCloud modèle 2To : stockant uniquement les scripts
- Un onduleur APC alimentant la box, son boîtier fibre (ONT), et les RPi.

6.11.1. Concernant la BBox

Toutes les box modernes ont par défaut tous les ports entrants fermés. La BBox a été configurée pour rediriger les flux entrants des ports 80 (http) et 443 (https) sur Raspi1, le port 1194 pour OpenVPN se dirigeait sur Raspi2.

Ainsi, le paramétrage effectué forçait l'accès au questionnaire via une connexion chiffrée. Une demande d'accès par un ordinateur extérieur au site <http://maguypaul.ovh> (HTTP non chiffré) force une redirection vers l'adresse <https://maguypaul.ovh> (remarquez le **HTTPS**).

6.11.2. Le Raspberry Pi et Raspbian

Le support matériel du questionnaire a été un RPi modèle 3B : un nano-ordinateur mono carte (SBC ou *Single Board Computer*).

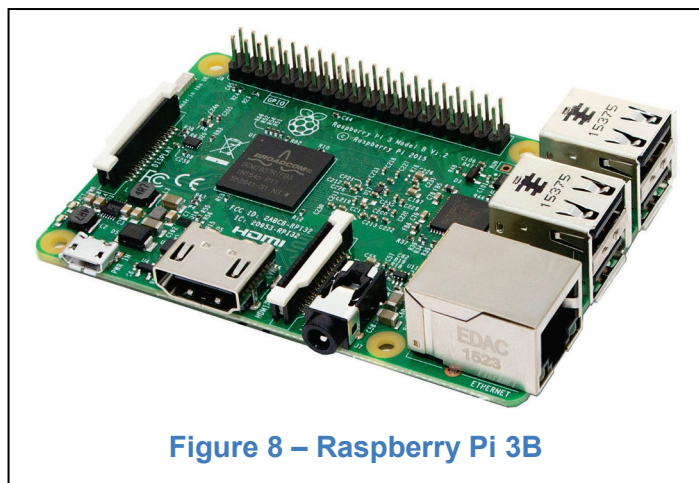


Figure 8 – Raspberry Pi 3B

Cet ordinateur est peu coûteux à l'achat : 35€ seul, moins de 100€ si l'on compte carte SD, boîtier et adaptateur secteur. Cet équipement a été réutilisé pour d'autres usages quotidiens une fois le travail de thèse effectué : il ne s'agit donc pas de frais perdus comme le sont les services payants d'hébergement de questionnaires en ligne.

Il est également économique à l'usage car ne consommant que l'équivalent de 5€ d'électricité par an si allumé 24h/24. Il est également compact : sa

surface est approximativement celle d'une carte de crédit. Enfin, il est suffisamment puissant pour de l'hébergement web simple.

Ce sont des ordinateurs à visée éducative et d'apprentissage, aux composants peu onéreux. Le cœur du système est une puce intégrant de multiples composants en un : il s'agit d'un [“System on a Chip” \(SoC\)](#), comparable en architecture à ceux utilisés dans les smartphones modernes.

En termes de puissance de calcul, un RPi 3B est grossièrement comparable à un ordinateur de bureau de l'année 2007. En revanche, les coûts à l'usage, encombrement et consommation électrique sont nettement inférieurs à ceux d'un ordinateur “standard” de pleine taille.

Au total, deux exemplaires identiques de RPi ont servi pour le questionnaire : l'un acheté spécifiquement pour le questionnaire, l'autre étant déjà en place pour d'autres tâches personnelles. Ce dernier a pu servir en parallèle pour la duplication de la base de données en temps réel, et donc permettre une sécurisation supplémentaire des données du questionnaire.

Le SoC des RPi 3B est une puce Broadcom BCM2837 64 bit, à quatre cœurs ARM Cortex-A53 à 1,2 GHz. Il dispose de 1Go de LPDDR2, et utilisait comme support de stockage principal une carte microSD de 32Go.

Les RPi ont fonctionné avec le système d'exploitation Raspbian. Il s'agit d'une distribution Linux basée sur un noyau Debian, spécialement modifié pour l'architecture ARM des RPi. Raspbian est fourni par la [Raspberry Pi Foundation](#).

Plus particulièrement, la version ‘jessie’ de Raspbian a été utilisée. Elle est téléchargeable via une [page d'archives](#), et installable sur une carte micro SD via un programme de gravure d'image comme [Etcher](#).

Lors des premières expérimentations du thésard, avant la réalisation du questionnaire, plusieurs réinstallations successives de Raspbian ont été faites sur le premier RPi.

Des scripts permettant d'automatiser les procédures d'installation et de réglages ont été écrites, et sont reportés ici en police Courier New.

Ils pourraient aider d'éventuels autres thésards à suivre une méthodologie proche de la nôtre. Notez cependant que ces scripts ne sont pas forcément adaptés à l'installation d'une version plus récente de Limesurvey ou de Raspbian, mais ils pourront servir de base pour cela.

Un script de réinstallation rapide nommé `raspiandefault.sh` a été écrit pour régler Raspbian afin de pouvoir faire fonctionner Limesurvey :

```
#!/bin/bash
# Script Name: pi/scripts/raspiandefault.sh
# Author: Jay + Paul
# Date : 01.06.2018
#
# Description: The following script sets things up after a new raspbian jessie
install, to get ready for a LAMP + letsencrypt docker
# Run Information: This script is run manually.

clear
echo "*** LET'S GO BABY ***"
# Update the Pi before anything else
echo "----- Updating/Upgrade Pi -----"
sudo apt update -qq
sudo apt upgrade -y -qq
echo "--- Installing essentials ---"
#ntp and nfs-common are built-in raspbian jessie
#sudo service ntp start
sudo apt install htop tmux nload etherwake exfat-fuse ntp sysbench -y -qq
#ensuring correct Internet time setting
sudo service ntp stop && sudo ntpd -gq && sudo service ntp start
### Insert @boot fixes lines in /etc/rc.local
# Fixes ARMfreq -> 1200MHz, rpcbind@boot, apache2@boot
echo "--- Fixing rpcbind at boot ---"
sudo su -c "sed -i '/^exit 0$/i # Fixing freqs at boot: ondemand\nsudo sh -c\n\"echo ondemand > /sys/devices/system/cpu/cpu0/cpufreq/scaling_governor\"'\n\n#\nFixing Apache2 on boot\nsudo systemctl daemon-reload\n\n#\nFixing rpcbind on\nboot\nsudo rpcbind service restart\n' /etc/rc.local"

# Injecting WDMycloud mount points in fstab
echo "--- Applying rpcbindfix now to allow nfs mounts ---"
sudo systemctl enable rpcbind
sudo service rpcbind restart
echo "--- Injecting WD mountpoints in fstab ---"
sudo su -c "echo '192.168.1.50:/mnt/HD/HD_a2/deluge /download/deluge nfs\nrw,nosuid,dev,exec,auto,nouser,async,hard,noatime,intr 0 2' >> /etc/fstab"
sudo su -c "echo '192.168.1.50:/mnt/HD/HD_a2/pi /mnt/wdapi nfs defaults 0 2' >>\n/etc/fstab"

# Disabling unused wireless
echo "----- Saving power -----"
sudo su -c "echo \"dtoverlay=pi3-disable-bt\ndtoverlay=pi3-disable-wifi\" >> /boot/config.txt"

# Disable BT Modem service
sudo systemctl disable hciuart

# Blacklist to forbid further installs
```

```

sudo su -c "echo \"#wifi
blacklist brcmfmac
blacklist brcmutil

#bt
blacklist btbcm
blacklist hci_uart

#sound
#blacklist intel_powerclamp
#blacklist pcspkr
#blacklist soundcore
#blacklist snd
#blacklist snd_hda_codec_hdmi
#blacklist snd_hda_codec_realtek
#blacklist snd_hda_codec_generic
#blacklist snd_hda_intel
#blacklist snd_hda_codec
#blacklist snd_hda_core
#blacklist snd_hwdep
#blacklist snd_intel_sst_acpi
#blacklist snd_intel_sst_core
#blacklist snd_soc_rt5670
#blacklist snd_soc_sst_mfld_platform
#blacklist snd_soc_rl6231
#blacklist snd_soc_sst_match
#blacklist snd_soc_core
#blacklist snd_compress
#blacklist snd_pcm
#blacklist snd_timer\" >> /etc/modprobe.d/raspi-blacklist.conf"

# Adding aliases commands to bash profile
echo "--- Adding aliases commands to bash profile ---"
echo "alias nbp=\"nano ~/.bash_profile && source ~/.bash_profile\"
alias ll=\"ls -alH --color=auto\"
alias temp=\"vcgencmd measure_temp\"
alias cls=\"clear\"
alias dps=\"sudo docker ps -a\"
alias dstop=\"sudo docker stop $(dps -q | tr '\n' ' ')\n\"
alias dstart=\"sudo docker start $(dps -q | tr '\n' ' ')\n\" >> ~/.bash_profile

#####
# Installing docker #
#####
echo "--- Installing Docker ---"
sudo apt install -y apt-transport-https ca-certificates curl gnupg2 software-
properties-common
curl -fsSL https://download.docker.com/linux/debian/gpg | sudo apt-key add -
echo "deb [arch=armhf] https://download.docker.com/linux/debian $(lsb_release -
cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list

sudo apt update -qq
sudo apt install -y -qq docker-ce
echo "\n\n\n*** SCRIPT OVER ***
You will next need to reboot, then execute raspil.sh.\n"
read -p "*** Press enter to reboot ***"
sudo reboot

```

S'ensuit le script raspil.sh à lancer après redémarrage :

```

#!/bin/bash
# Script Name: pi/scripts/raspil.sh
# Author: Jay + Paul

```

```

# Date : 01.06.2018
# Description: The following script is to be used on raspil.
#             sets up letsencrypt + nginx-proxy dockers,
# Run Information: This script is run manually.
# WARNING: It must be run AFTER raspbiandefault.sh

clear
sudo mkdir /srv/letsencrypt
sudo mkdir /srv/nginx-proxy
sudo mkdir /srv/docker_conf

echo "----- Setting up portainer docker -----"
sudo docker run -d -p 9000:9000 --name portainer --restart always -v
/var/run/docker.sock:/var/run/docker.sock -v /srv/docker_conf/portainer:/data
portainer/portainer

echo "----- Setting up nginx-proxy docker -----"
sudo docker run -d --restart unless-stopped -p 443:443 --name nginx-proxy -v
/srv/nginx-proxy/ssl.conf:/etc/nginx/ssl.conf:ro -v /srv/nginx-
proxy/default.conf:/etc/nginx/conf.d/default.conf:ro -v
/srv/letsencrypt/etc/letsencrypt/:/etc/nginx/certs:z -v
/srv/letsencrypt/nginx/dhparams.pem:/etc/nginx/dhparam.pem:ro nginx

echo "----- Setting-up lsioarmhf/letsencrypt docker -----"
sudo docker run -d --cap-add=NET_ADMIN --name=letsencrypt --restart unless-
stopped -v /srv/letsencrypt:/config:z -e PGID=1000 -e PUID=1000 -e EMAIL=[ADRESSE
PERSO]@laposte.net -e URL=maguypaul.ovh -e SUBDOMAINS=www -e VALIDATION=http -p
80:80 -e TZ=Europe/Paris lsioarmhf/letsencrypt
echo "\n\n*** SCRIPT END ***\n\n"

```

Avec ces scripts, en cas de dysfonctionnement grave d'un Raspberry Pi et si la carte SD venait à ne plus fonctionner, une réinstallation complète n'aurait duré qu'une dizaine de minutes. Un tel scénario ne s'est pas produit, mais encore une fois ces informations pourraient intéresser de futurs thésards.

6.11.3. Installation de Limesurvey sur Raspbian

La version de Limesurvey utilisée était la 2.64, disponible en 2017 lors de la création des premiers questionnaires du thésard.

En se basant sur un tutoriel [\(64\)](#) qui mentionnait les commandes à exécuter, un script d'installation rapide nommé limeinstall.sh a été réalisé :

```

#!/bin/bash
# Script Name: pi/scripts/limeinstall.sh
# Author: Jay + Paul
# Date : 01.06.2018
#
# Description: The following script installs limesurvey v2.64.0+170307
# Run Information: This script is run manually.

clear
echo "*** Installing Limesurvey v2.64.0+170307 ***\n"
echo "--- Installing Apache2 ---"
sudo apt install apache2 -y -qq
cd /var/www/html
sudo chmod 775 .

echo "--- Installing MySQL server ---"
sudo apt install mysql-server -y
# mySQL passwd asked here

```

```

echo "--- Installing php5 and mods ---"
sudo apt install php5 libapache2-mod-php5 php5-mysqldb php5-gd php5-ldap php5-
imap -y -qq
# The simple "php5-mysql" package is NOT compatible with "php5-mysqldb" and not
to be installed, it seems
# sudo apt install php5-mysql -y -qq

echo "--- Installing PHPMyAdmin ---"
sudo apt install phpmyadmin -y -qq

echo "
--- Downloading Limesurvey v2.64.0+170307 ---"
# Editing to be done in this section to download the latest version of LS
cd ~/Downloads
wget https://github.com/LimeSurvey/LimeSurvey/archive/2.64.0+170307.zip
sudo mv 2.64.0+170307.zip /var/www/html
cd /var/www/html
sudo unzip 2.64.0+170307.zip
sudo mv LimeSurvey-2.64.0-170307/ limesurvey
#sudo rm 2.64.0+170307.zip

sudo chown -R www-data:www-data /var/www/html/limesurvey
sudo chmod -R 775 /var/www/html/limesurvey

echo "--- Restarting apache2 ---"
sudo systemctl daemon-reload
sudo service apache2 restart

echo "\nGo to Pi IP/limesurvey
import missing templates
and let's go!
\n
*** SCRIPT OVER ***

If your installation is complex, consider changing Apache2 ports.
/etc/apache2/ports.conf -> listen 80? or 808x?"

```

6.11.4. Sauvegarde Limesurvey Rebatel

La méthode de sauvegarde des données de Limesurvey était un “*dump*” quotidien de sa base de données MySQL. Signifiant “déversement”, ce procédé consiste en la création d’un fichier brut décrivant l’intégralité des paramètres et contenus d’une base de données.

Ce “dump” était lancé automatique via le “cron” : le programme qui permet aux utilisateurs des systèmes Unix d’exécuter automatiquement des scripts, des commandes ou des logiciels à une date et une heure spécifiées à l’avance. Le cron est accessible via la commande `crontab -e`. La ligne ajoutée était :

```
0 4 * * * /mnt/wdpi/scripts/cron/backupmysql.sh
```

ce qui se traduit par : lancer le script `backupmysql.sh` tous les jours à 4 heures du matin.

Le script bash permettant cette opération par Raspi 1 était :

```
#!/bin/bash
# Script Name: pi/scripts/backupmysql.sh
# Original author: Jay

```



```

# Modified by Paul
# Date : 06-2018
#
# Description: Dumps all MySQL DBs except those in DATABASES_TO_EXCLUDE into a
.sql file
#
#           If previous' day file = this day's file -> rm previous day file.
#           If 1st day of the month, creates a .tb2 file with previous's month
.sql files.
# Warning : NEEDS a mysql user called 'backup' with appropriate permissions on
tables dumped.
# Run Information: This script is to be run by crontab.
# Optional arguments : -delete authorizes deletion of old .sql files after last
month's taring.

clear
DATE=`date +%Y %m %d`
DATE_YESTERDAY=`date +%Y %m %d --date="-1 day"`
FILE=mysqldump_`DATE`.sql
FILE_YESTERDAY=mysqldump_`DATE_YESTERDAY`.sql
PATH_BACKUP="/mnt/wdpi/backups/mysql"

# Running DB exclusion routine. Add DBs to exclude here.
DATABASES_TO_EXCLUDE="phpmyadmin"
EXCLUSION_LIST=""information schema','performance_schema','mysql'"
for DB in `echo "${DATABASES_TO_EXCLUDE}"`
do
    EXCLUSION_LIST="${EXCLUSION_LIST},'${DB}'"
done

SQLSTMT="SELECT schema_name FROM information_schema.schemata WHERE schema_name
NOT IN (${EXCLUSION_LIST})"
MYSQLDUMP_DATABASES="--databases"
for DB in `mysql -u backup -ANe"${SQLSTMT}"`
do
    MYSQLDUMP_DATABASES="${MYSQLDUMP_DATABASES} ${DB}"
done

# Dump options here
MYSQLDUMP_OPTIONS="--routines --triggers"

echo -e "--- Dumping MySQL Databases ---\n\nExcluded DataBases are:
$EXCLUSION_LIST\n\nDumped DataBases are: $MYSQLDUMP_DATABASES\n\nDestination
file: $PATH_BACKUP/$FILE\n\n"
## old simple command (was only for MaguyDB): "mysqldump -u backup --databases
MaguyDB > $PATH_BACKUP/$FILE"
mysqldump -u backup ${MYSQLDUMP_OPTIONS} ${MYSQLDUMP_DATABASES} >
$PATH_BACKUP/$FILE

# Check if yesterday's file is the same
if [ -f $FILE_YESTERDAY ]
then
    DIFF=$(diff -y --suppress-common-lines $PATH_BACKUP/$FILE
$PATH_BACKUP/$FILE_YESTERDAY | wc -l -1)
    if [ $DIFF -eq "0" ]
    then
        echo -e "$FILE is strictly identical to $FILE_YESTERDAY\n => Deleting
$FILE_YESTERDAY"
        rm $PATH_BACKUP/$FILE_YESTERDAY
    else echo "File was different -> no deletion."
    fi
fi

if [ `date +%d` -eq 1 ]
then LASTMONTH=mysqldump_`date +%Y %m --date="-1 month"`
# tars last month's dumps in one file in archive folder
tar cfj $PATH_BACKUP/archives/mysqldumps_`date +%Y %m`.tb2

```

```

$PATH_BACKUP/$LASTMONTH\_*\.sql

if ! [ -z $1 ]
then
# Deletes only if -delete is set.
if [ $1 = "-delete" ]
then echo "Rming old files authorized" && rm
$PATH_BACKUP/$LASTMONTH\_*\.sql
else echo "Incorrect argument."
fi
else echo "Old files left in place (launch with -delete to change this)."

```

Le résultat du dump était initialement stocké localement sur la carte SD, et une alerte était envoyée par mail au thésard pour lui signaler la réalisation de l'opération.

En l'état, elles étaient tout à fait ré installables sur une installation MySQL tierce, afin de pouvoir les restaurer ou les lire via une autre installation Limesurvey.

Ensuite, une installation en Docker du programme Duplicati était chargée de créer une sauvegarde incrémentale chiffrée du dump. Cette sauvegarde était envoyée sur un compte Google Drive quotidiennement vers 4h du matin.

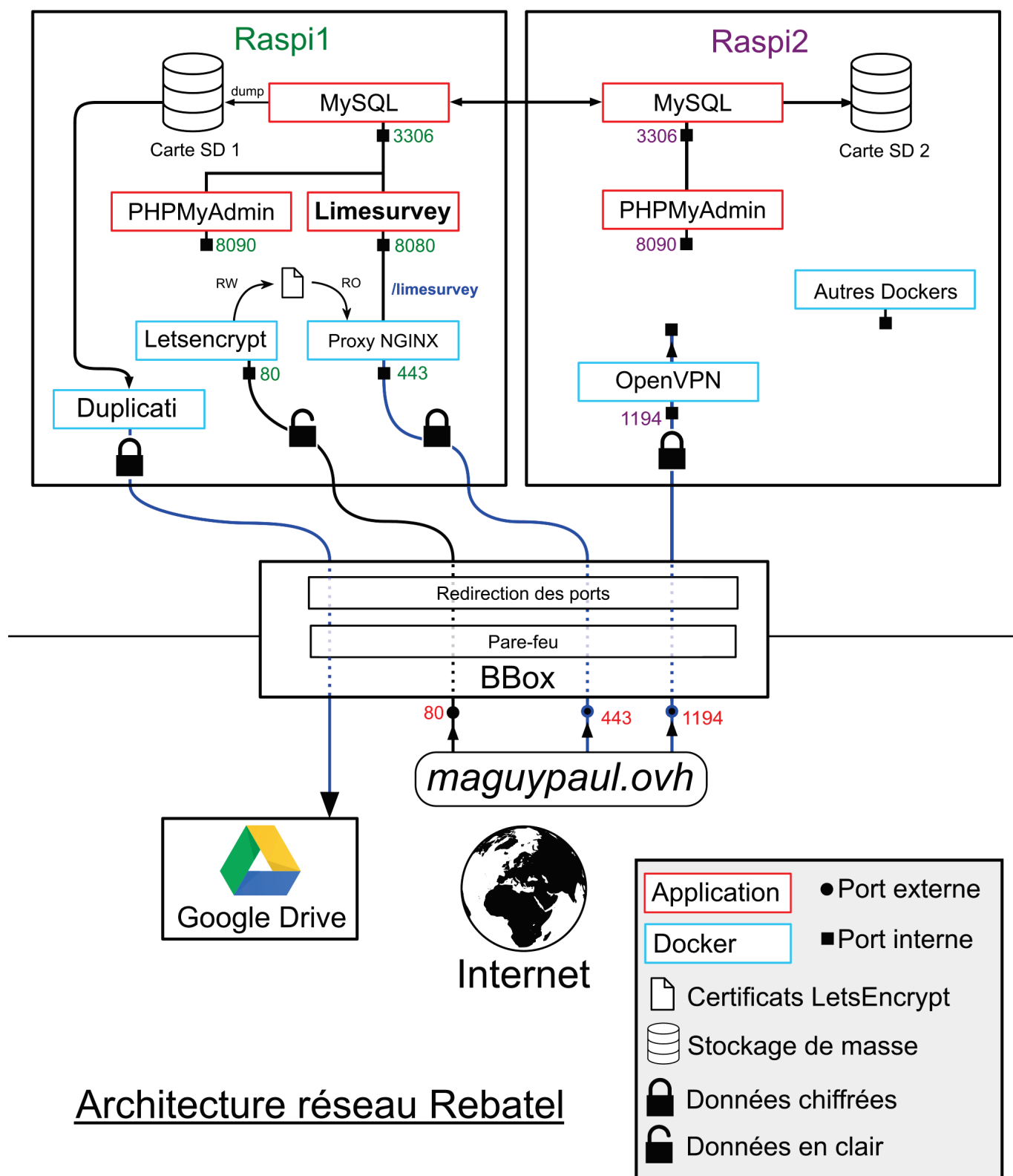
Le chiffrement employé était la méthode AES-256, dont la clé était un mot de passe de 12 caractères, basé sur une phrase. Ainsi, même si la sauvegarde est stockée chez Google, il est extrêmement improbable que ces données soient exploitables par Google, car les données sont chiffrées de façon robuste.

Dans notre situation particulière, la quantité totale de données sauvegardées (en comptant les sauvegardes incrémentales) n'a pas excédée les 50Mo : c'est en 2018 une quantité de données relativement faible.

Si l'on avait employé une connexion Internet haut-débit type ADSL, il aurait fallu au maximum plusieurs minutes pour l'intégralité de la sauvegarde, et moins dans le cas de la méthode incrémentale. Dans notre cas, l'accès fibré a permis des exports itératifs en quelques secondes.

6.11.5. Schéma récapitulatif

Figure 9 - Architecture réseau Rebatel



Architecture réseau Rebatel

Nota : le NAS WD MyCloud, porteur des scripts d'installation et de sauvegarde, n'est pas représenté. Ces scripts n'ont pas été stockés à même le RPi car le contenu de sa carte SD, hormis la base MySQL, n'était pas sauvegardé ou dupliqué, ce qui était le cas du WD MyCloud.

6.11.6. Incidents techniques “Rebatel”

Durant la période de 5 mois de fonctionnement, deux incidents techniques mineurs se sont produits :

- Péremption des certificats LetsEncrypt présentés par le proxy NGINX :

Cet incident a empêché l'accès au serveur “Rebatel” pendant la période du début du mois de septembre, rendant impossible l'accès au questionnaire sur ce serveur pendant une journée.

La cause du problème était le mode de fonctionnement du proxy NGINX d'accueil sur le port 443.

Pour comprendre, expliquons le fonctionnement de LetsEncrypt : le certificat nécessaire à l'authentification et donc l'accès en HTTPS est par défaut renouvelé tous les 3 mois grâce au Docker LetsEncrypt. Les fichiers de certificat sont disponibles en permanence aux autres composants systèmes, dont le docker du proxy NGINX d'accueil. Cependant, cette instance NGINX, avec certains réglages restés par défaut, ne se relançait ni ne rechargeait les nouveaux certificats une fois ceux-ci renouvelés. Ainsi, malgré le renouvellement effectif des fichiers du certificat, présents sur le Raspi1, ceux-ci n'étaient pas chargés par le proxy périodiquement. Ainsi, un répondant pouvait se voir présenter une page lui indiquant une connexion non sécurisée à cause d'un certificat périmé. La correction du problème a consisté en un redémarrage du proxy NGINX, qui une fois relancé a chargé les certificats à jour.

Pour éviter ce problème, il aurait fallu paramétrer un redémarrage périodique du Docker porteur du proxy NGINX, par une ligne supplémentaire dans le cron type `sudo docker restart nginx-proxy` tous les mois par exemple. Etant donné la rapidité de lancement de ce Docker, la conséquence en serait au pire, si sollicitation au moment exact de son redémarrage, un accès impossible une seule fois. Lancer un redémarrage en heures creuses limiterait ce risque déjà très faible.

- Plantage du serveur MySQL secondaire sur le 2ème Raspberry Pi :

Suite à redémarrage du Raspi 2 (erreur humaine du thésard, commande lancée dans le terminal et regrettée immédiatement ensuite), la synchronisation des deux instances MySQL en master-master a été rompue pendant plusieurs heures, empêchant la collecte de réponses sur le serveur “Rebatel”. Etant donné la prise de conscience immédiate de l'erreur, un appel au secours envoyé à M. MICHEL Jérôme, bien plus expérimenté en synchronisation de bases de données MySQL que le thésard, a permis de réparer la synchronisation dans les heures qui ont suivi.

6.12. Installation redondante “MiCloud”

L'installation redondante a été faite sur un serveur personnel physiquement situé dans la ville d'Antibes appartenant à M. MICHEL Jérôme, accessible via l'URL “micloud.ovh/limesurvey”.

Elle a utilisé les mêmes logiciels que ceux employés sur l'installation Rebatel : Limesurvey, base MySQL. Le système d'exploitation était proche, un Debian en version “stretch”.

Le support d'hébergement a été un Mac Pro modèle 6.1 équipé d'un processeur Intel® Xeon® E5-1680 v2 (8 cœurs, 16 threads) architecture x64, et de 64Go de Ram [ECC](#). Le stockage des données

est réalisé sur un NAS rattaché au réseau local. Une machine de cette classe peut être utilisée en tant que station de travail ou serveur d'entreprise, et ce encore en 2019.

6.12.1. Différences par rapport à l'installation Rebatel

Contrairement à l'installation "Rebatel", "Micloud" a fonctionné à l'intérieur d'une [machine virtuelle](#). Un serveur d'entreprise est typiquement capable de supporter plusieurs machines virtuelles simultanément, comme explicité sur le site de Microsoft :

"Il est possible d'exécuter plusieurs machines virtuelles simultanément sur un même ordinateur physique. Pour les serveurs, les divers systèmes d'exploitation fonctionnent côte à côte, avec un composant logiciel appelé hyperviseur pour les gérer, alors que les ordinateurs de bureau classiques n'utilisent qu'un seul système d'exploitation pour exécuter d'autres systèmes d'exploitation dans des fenêtres de programme qui leur sont propres. Chaque machine virtuelle fournit son propre matériel virtuel, à savoir les processeurs, la mémoire, les disques durs, les interfaces réseau et les autres périphériques nécessaires. Le matériel virtuel est ensuite mappé au matériel réel sur la machine physique, ce qui permet de réaliser des économies en réduisant le besoin de disposer de systèmes matériels physiques, ainsi que les coûts de maintenance associés, tout en réduisant la demande en alimentation et refroidissement." [\(65\)](#)

Dans ce cas précis, l'hyperviseur était une machine VMware. Les ressources allouées à la machine virtuelle hébergeant l'installation de Limesurvey avait comme ressources allouées un cœur CPU et 1Go de Ram. La Ram était une mémoire de type [ECC](#), donc moins susceptible aux erreurs que la Ram standard DDR2 des RPi.

6.12.2. Sauvegarde Limesurvey Micloud

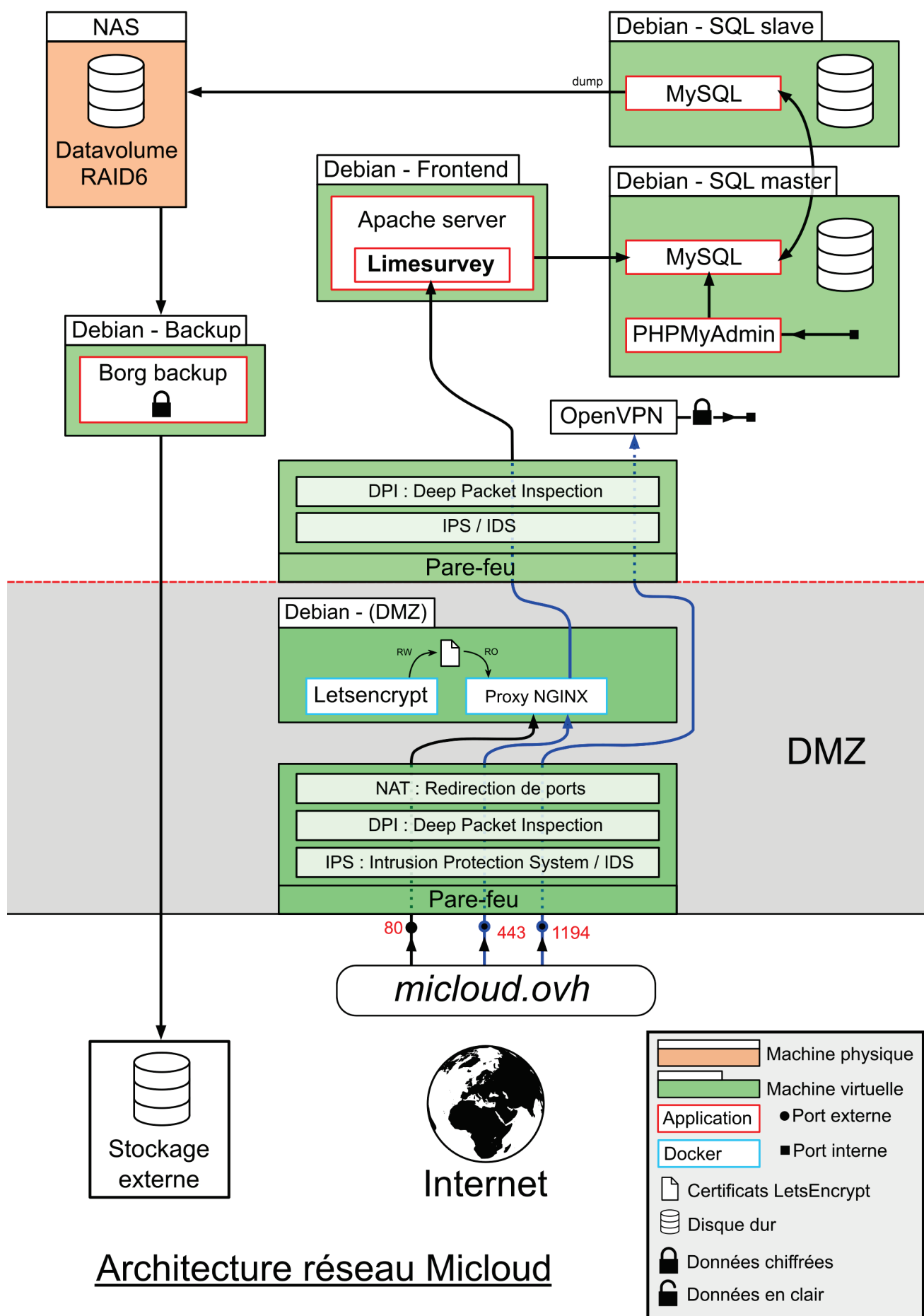
Les données hébergées par l'installation Limesurvey au sein d'une machine virtuelle étaient stockées de façon incrémentale sur un NAS contenant plus de 3 disques durs configurés en [RAID 6](#), au rythme de :

- un [instantané](#) par jour, conservé pendant 2 mois
- une sauvegarde complète une fois par jour, conservée pendant un mois
- une sauvegarde par semaine conservée pour les périodes entre 1 et 6 mois
- une sauvegarde par mois pour les périodes au-delà de 6 mois.

Les données de la base de données du questionnaire étaient sauvegardées séparément selon le même principe.

6.12.3. Schéma récapitulatif

Figure 10 - Architecture réseau Micloud



6.13. Résultats détaillés

Tableau 56 : Configuration de types de supports (physiques et en *cloud*) détaillée

Les configurations ayant une fréquence inférieure à 1%, non présentées dans les résultats, sont ici surlignées en gris clair.

Codage des supports : supports locaux en majuscules, supports externalisés en minuscules.

H/h = Disque dur standard, S/s = SSD, N/n = NAS, O/o = Optique, U/u = clé USB.

Configurations de supports physiques	Sous-total	+ progiciel <i>cloud</i> seul	+ export <i>cloud</i> tiers seul	+ les deux	Total	Pourcentage
h	116	6	4	1	127	20,85%
Aucun support physique	12	70	14	2	98	16,09%
Hh	71		3	1	75	12,32%
uH	37	2	2		41	6,73%
U	34	3	2	1	40	6,57%
Uu	27	2	2		31	5,09%
n	15	10	3	2	30	4,93%
HN	20		4		24	3,94%
HS	14		4		18	2,96%
UuH	11	2			13	2,13%
HhN	8		1		9	1,48%
HSN	8				8	1,31%
HhSs	6		2		8	1,31%
Un	5		1		6	0,99%
UuHh	5				5	0,82%
UHn	5				5	0,82%
UuN	4		1		5	0,82%
HhS	3	1	1		5	0,82%
UHh	5				5	0,82%
Nn	3		1		4	0,66%
uHS	4				4	0,66%
HhNn	3		1		4	0,66%
Oh	2		1		3	0,49%
Uuhn	3				3	0,49%
uHhs	3				3	0,49%
hNn	3				3	0,49%
oN	2				2	0,33%
HhSsN	2				2	0,33%
HhSN	2				2	0,33%
uHhN	2				2	0,33%
UuHhSs	2				2	0,33%
Oo	1	1			2	0,33%
oHh	2				2	0,33%
UHhSs	2				2	0,33%
OoHh	2				2	0,33%
OHN	1				1	0,16%
UuOo	1				1	0,16%
UNn			1		1	0,16%
uoHN	1				1	0,16%
UuOoHhsNn	1				1	0,16%
HSNn	1				1	0,16%
UuHS	1				1	0,16%
O		1			1	0,16%
UuNn	1				1	0,16%
OoH			1		1	0,16%
UuOoHh	1				1	0,16%
UuHhSN		1			1	0,16%
uo	1				1	0,16%
uHSN	1				1	0,16%
Total général	454	99	49	7	609	100%

Tableau 57 : Cas des utilisateurs de progiciel *cloud* et *cloud* tiers détaillé

Configurations de supports physiques	Progiciel <i>cloud</i> seul	Export <i>cloud</i> tiers seul	Les deux	Total	Pourcentage des 609 répondants
(aucun)	70	14	2	86	14,12%
n	10	3	2	15	2,46%
h	6	4	1	11	1,81%
U	3	2	1	6	0,99%
Uu	2	2		4	0,66%
HN		4		4	0,66%
uH	2	2		4	0,66%
HS		4		4	0,66%
Hh		3	1	4	0,66%
HhS	1	1		2	0,33%
HhSs		2		2	0,33%
UuH	2			2	0,33%
HhN		1		1	0,16%
Un		1		1	0,16%
HhNn		1		1	0,16%
UNn		1		1	0,16%
Oh		1		1	0,16%
OoH		1		1	0,16%
Oo	1			1	0,16%
UuHhSN	1			1	0,16%
UuN		1		1	0,16%
Nn		1		1	0,16%
O	1			1	0,16%
Total général	99	49	7	155	25,45% (de 609)

Tableau 58 : Résultats ANOVA multivariée détaillés

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
âge	3	133.0	44.34	8.535	1.49e-05 ***
sexe	1	112.3	112.31	21.617	4.14e-06 ***
groupe	1	19.0	18.99	3.655	0.05639
ancienneté	3	45.5	15.17	2.920	0.03356 *
âge:sexe	3	2.0	0.66	0.126	0.94456
âge:groupe	3	16.9	5.62	1.081	0.35639
sexe:groupe	1	4.5	4.53	0.873	0.35056
âge:ancienneté	3	23.8	7.93	1.526	0.20674
sexe:ancienneté	3	4.2	1.39	0.268	0.84847
groupe:ancienneté	3	10.2	3.41	0.656	0.57941
âge:sexe:groupe	2	5.5	2.73	0.525	0.59201
âge:sexe:ancienneté	3	59.8	19.94	3.838	0.00971
âge:groupe:ancienneté	3	12.7	4.23	0.815	0.48607
sexe:groupe:ancienneté	3	15.8	5.28	1.016	0.38491
âge:sexe:groupe:ancienneté	2	13.0	6.48	1.248	0.28792
Residuals	571	2966.4	5.20		

On ne retient pas les tests d'interactions entre les facteurs d'étude, car ils ne sont pas interprétables du fait de leur trop grand nombre qui augmente le risque alpha.

6.14. Références bibliographiques

1. mssante.fr. MSSanté - découvrir (industriels) - <https://mailiz.mssante.fr/is/decouvrir>
2. Réseau privé virtuel - [https://fr.wikipedia.org/w/index.php?title=Réseau privé virtuel](https://fr.wikipedia.org/w/index.php?title=Réseau_privé_virtuel)
3. IPSOS France. Les médecins à l'ère du numérique - <http://www.ipsos.fr/decrypter-societe/2017-01-31-medecins-l-ere-numerique>
4. Nelson S. Pro Data Backup and Recovery, Berkeley, CA: Apress - <http://link.springer.com/10.1007/978-1-4302-2663-5>
5. Bami G. Protection des données patient informatisées en médecine générale: étude transversale, département du Val-d'Oise, septembre-décembre 2007 [Thèse d'exercice]. [France]: Université Paris Diderot - Paris 7. UFR de médecine; 2009. - <http://www.sudoc.fr/13793260X>
6. Cheriet M. Sécurité des données médicales informatiques en cabinet de médecine générale [Thèse d'exercice]. [France]: Université Pierre et Marie Curie (Paris). UFR de médecine Pierre et Marie Curie; 2012. - <http://www.sudoc.fr/164389296>
7. Staechel R, Université Louis Pasteur Strasbourg. Sécurité des données informatiques : à propos de la sauvegarde des données de santé au cabinet de médecine générale : enquête auprès des praticiens du Haut-Rhin - avril 2012. [Thèse d'exercice]. 2012. - <http://www.sudoc.fr/16489862X>
8. sesam-vitale.fr. Notre histoire [Internet]. - <http://www.sesam-vitale.fr/web/guest/notre-histoire>
9. Walley T, Mantgani A. The UK General Practice Research Database. THE LANCET. 1997;350:3.
10. Logiciels médicaux [Internet]. - <http://comparatif-logiciels-medicaux.fr/logiciels-bureau>
11. Larousse. Expressions : nuage [Internet]. - <https://www.larousse.fr/dictionnaires/francais/nuage/55167>
12. Téléphonie mobile. In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=Téléphonie mobile](https://fr.wikipedia.org/w/index.php?title=Téléphonie_mobile)
13. eduscol.education.fr. Sécurité des systèmes d'information : de la gestion des risques à la confiance numérique [Internet]. eduscol. - http://eduscol.education.fr/ecogest/si/SSI/risk_conf
14. Risque (informatique). In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=Risque \(informatique\)](https://fr.wikipedia.org/w/index.php?title=Risque_(informatique))
15. missionmode.com. Lies, #@!* Lies and Data Loss Statistics & IT Disaster Recovery Stats [Internet]. MissionMode. 2012 - <https://www.missionmode.com/lies-lies-and-data-loss-statistics/>
16. abnewswire.com. Data Loss Prevention Market 2018 Global Analysis, Growth, Size, Share, Trends, Forecast to 2022 [Internet]. - http://www.abnewswire.com/pressreleases/data-loss-prevention-market-2018-global-analysis-growth-size-share-trends-forecast-to-2022_280043.html
17. Vättö K. The Truth About SSD Data Retention [Internet] - <https://www.anandtech.com/show/9248/the-truth-about-ssd-data-retention>
18. Plan SSI | Accompagnement Cybersécurité des Structures de Santé [Internet] - <https://cyberveille-sante.gouv.fr/plan-ssi>
19. France B nationale de. BnF - La numérisation de masse à la BnF [Internet] - http://www.bnf.fr/fr/collections_et_services/anx_bib_num/a.numerisation_masse_bnf.htm
20. StorageCraft IT News. Data Loss Statistics - Infographic [Internet]. StorageCraft Technology Corporation. 2017 - <https://blog.storagecraft.com/data-loss-statistics-infographic/>
21. Bremme L. Définition : Qu'est-ce que le Big Data ? [Internet]. LeBigData.fr. - <https://www.lebigdata.fr/definition-big-data>
22. CNIL. Affaire Cambridge Analytica / Facebook [Internet]. - <https://www.cnil.fr/fr/affaire-cambridge-analytica-facebook>
23. Hern A. Google will stop scanning content of personal emails [Internet]. the Guardian. 2017 - <http://www.theguardian.com/technology/2017/jun/26/google-will-stop-scanning-content-of-personal-emails>

24. modernhealthcare.com. Data Points: Get facts and figures on the state of healthcare IT. [Internet]. Modern Healthcare. - <http://www.modernhealthcare.com/article/20170121/MAGAZINE/301219987>
25. santé.lefigaro.com CJ. Les données de santé attirent les hackers [Internet]. 2015 - <http://sante.lefigaro.fr/actualite/2015/02/13/23393-donnees-sante-attirent-hackers>
26. Lee CH, Yoon H-J. Medical big data: promise and challenges. Kidney Res Clin Pract. mars 2017;36(1):3-11.
27. CNIL. Norme simplifiée NS-050 - Cabinet médical et paramédical [Internet]. - <https://www.cnil.fr/fr/declaration/ns-050-cabinet-medical-et-paramedical>
28. TEIL E, ANNE D. Le Règlement Général sur la Protection des Données. URPS AuRA - Trait Union. 1 oct 2018;(10):4-5.
29. CNIL. Le registre des activités de traitement [Internet]. - <https://www.cnil.fr/fr/RGDP-le-registre-des-activites-de-traitement>
30. CNIL. RGPD - Chapitre VIII - Voies de recours, responsabilité et sanctions [Internet]. - <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre8#Article84>
31. CNIL. OPTICAL CENTER : sanction de 250.000€ pour une atteinte à la sécurité des données des clients du site internet www.optical-center.fr [Internet]. - <https://www.cnil.fr/fr/optical-center-sanction-de-250000eu-pour-une-atteinte-la-securite-des-donnees-des-clients-du-site>
32. Séneca H. Exame Informática | CNPD: Hospital do Barreiro multado em 400 mil euros por permitir acessos indevidos a processos clínicos [Internet]. Exame Informática. - <http://exameinformatica.sapo.pt/noticias/mercados/2018-10-19-CNPD-Hospital-do-Barreiro-multado-em-400-mil-euros-por-permitir-acessos-indevidos-a-processos-clinicos>
33. CNIL. RGPD : se préparer en 6 étapes [Internet]. - <https://www.cnil.fr/fr/principes-cles/rgpd-se-preparer-en-6-etapes>
34. CNIL. RGPD et professionnels de santé libéraux : ce que vous devez savoir [Internet]. - <https://www.cnil.fr/fr/rgpd-et-professionnels-de-sante-liberaux-ce-que-vous-devez-savoir>
35. CNIL. Pratiques abusives « Mise en conformité RGPD » : comment s'en prémunir avec la CNIL et la DGCCRF ? [Internet]. - <https://www.cnil.fr/fr/pratiques-abusives-mise-en-conformite-RGPD-CNIL-DGCCRF>
36. Agrément pour l'hébergement de données de santé [Internet]. - <http://esante.gouv.fr/labels-certifications/hds/agrement>
37. Serries G. Hébergeur de Données de Santé (HDS), pourquoi tout le monde s'y met ? [Internet]. ZDNet France. - <https://www.zdnet.fr/actualites/hebergeur-de-donnees-de-sante-hds-pourquoi-tout-le-monde-s-y-met-39880265.htm>
38. legifrance.gouv.fr. LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé [Internet]. 2016-41 janv 26, 2016. - <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031912641&categorieLien=id>
39. ASIP Santé. Certification des hébergeurs de données de santé [Internet]. esante.gouv.fr. - <http://esante.gouv.fr/labels-certifications/hds/certification-des-hebergeurs-de-donnees-de-sante>
40. CNOM. Atlas national [Internet]. - <https://www.conseil-national.medecin.fr/node/1476>
41. Cousin S. Installation, pourquoi les jeunes choisissent l'exercice en groupe [Internet]. Remede.org. - <http://www.remede.org/documents/installation-pourquoi-les-jeunes-choisissent-l-exercice-en-groupe.html>
42. CNOM. Dossiers médicaux : conservation et archivage [Internet]. - <https://www.conseil-national.medecin.fr/article/dossiers-medicaux-conservation-et-archivage-98>
43. Carnet de santé numérique : une bonne idée ? [Internet]. France Inter. - <https://www.franceinter.fr/emissions/le-telephone-sonne/le-telephone-sonne-12-novembre-2018>
44. Raby M, universfreebox.com. Scaleway (Iliad) annonce l'obtention de l'agrément « Hébergeur de Données de Santé » pour l'ensemble de ses data centers parisiens [Internet]. -

<https://www.universfreebox.com/article/48201/Scaleway-Iliad-annonce-l-obtention-de-l-agrement-Hebergeur-de-Donnees-de-Sante-pour-l-ensemble-de-ses-data-centers-parisiens>

45. Certes N. Microsoft certifié hébergeur de données de santé en France - Le Monde Informatique [Internet]. LeMondelInformatique. - <https://www.lemondeinformatique.fr/actualites/lire-microsoft-certifie-hebergeur-de-donnees-de-sante-en-france-73347.html>
46. Trély V, lequotidiendumedecin.fr. Les données de santé ont une valeur marchande, elles sont donc convoitées ! [Internet]. Le Quotidien du Médecin. - <https://www.lequotidiendumedecin.fr/opinions/article/2018/11/12/les-donnees-de-sante-ont-une-valeur-marchande-elles-sont-donc-convoitees- 862744>
47. Grandfather-Father-Son Backup. In: Wikipédia [Internet]. 2016 - https://fr.wikipedia.org/w/index.php?title=Grandfather-Father-Son_Backup&oldid=124887586
48. ASIP Santé. Editeurs : comment obtenir le label ? [Internet]. esante.gouv.fr. - <http://esante.gouv.fr/services/labellisation/editeurs-comment-obtenir-le-label>
49. Noyes K. Docker: A « Shipping Container » for Linux Code [Internet]. Linux.com | The source for Linux information. 2013 - <https://www.linux.com/news/docker-shipping-container-linux-code>
50. Docker (logiciel). In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=Docker_\(logiciel\)](https://fr.wikipedia.org/w/index.php?title=Docker_(logiciel))
51. Duplicati. In: Wikipedia [Internet]. 2019 - <https://en.wikipedia.org/w/index.php?title=Duplicati>
52. Mémoire à code correcteur d'erreurs. In: Wikipédia [Internet]. 2018 - https://fr.wikipedia.org/wiki/Mémoire_à_code_correcteur_d%27erreurs
53. Instantané (informatique). In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=Instantané_\(informatique\)](https://fr.wikipedia.org/w/index.php?title=Instantané_(informatique))
54. Historique (informatique). In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=Historique_\(informatique\)](https://fr.wikipedia.org/w/index.php?title=Historique_(informatique))
55. Open source. In: Wikipédia [Internet]. 2018 - https://fr.wikipedia.org/w/index.php?title=Open_source
56. RAID (informatique). In: Wikipédia [Internet]. 2018 - [https://fr.wikipedia.org/w/index.php?title=RAID_\(informatique\)](https://fr.wikipedia.org/w/index.php?title=RAID_(informatique))
57. Genma. Sauvegarde la règle des 3-2-1 [Internet]. - <https://blog.genma.fr/?Sauvegarde-la-regle-des-3-2-1>
58. Patrick M. 3-2-1: The Golden Rule for Backups [Internet]. Mirazon. 2015 - <http://www.mirazon.com/3-2-1-the-golden-rule-for-backups/>
59. Campbell M. Why 3-2-1 Backup Sucks | [Internet]. 2015 - <https://www.unitrends.com/blog/3-2-1-backup-sucks>
60. Système sur une puce. In: Wikipédia [Internet]. 2018 - https://fr.wikipedia.org/w/index.php?title=Système_sur_une_puce
61. Client léger. In: Wikipédia [Internet]. 2018 - https://fr.wikipedia.org/w/index.php?title=Client_léger
62. Larousse. Définitions : verbatim [Internet]. - <https://www.larousse.fr/dictionnaires/francais/verbatim/81475>
63. limesurvey.org. Stable release [Internet]. - <https://www.limesurvey.org/stable-release>
64. Leenes R. Limesurvey on a Raspberry Pi (3b) [Internet]. - <http://www.ronaldleenes.nl/stuff/limesurvey-on-raspberry/>
65. azure.microsoft.com. Machine virtuelle - Définition et fonctionnement [Internet]. - <https://azure.microsoft.com/fr-fr/overview/what-is-a-virtual-machine/>



Contexte et question de recherche :

Le RGPD a contribué à redonner de l'importance aux méthodes de protection des données médicales. En 2017-2018, comment les données médicales informatisées sont-elles protégées, sauvegardées et archivées par les médecins généralistes français ?

Méthodologie : Étude quantitative concernant les connaissances et pratiques des médecins généralistes français. Questionnaire en ligne auto-hébergé diffusé par sept URPS.

Résultats et discussion : 609 réponses sur six mois. Parmi les médecins répondants, 98% réalisaient des sauvegardes. Au cours des 5 dernières années, 18% avaient subi une perte de données, au moins en partie définitive pour la moitié d'entre eux. Seuls 25% des médecins ont déjà réalisé un test de restauration (unique pour 6% des médecins). La fiabilité des sauvegardes est largement perfectible. Les moyens de renforcement de résilience de sauvegarde sont sous-exploités. Les solutions progiciels et de stockage en *cloud* restent minoritaires.

Conclusions :

La sécurité des données médicales informatisées suppose une implication et une prise de conscience des généralistes. À défaut d'appliquer des corrections eux-mêmes à leur système informatique, ils doivent s'assurer que leur progiciel utilise des méthodes adaptées. Pour limiter les risques de perte de données dues à des mises à jour de logiciels ou à la rupture d'une chaîne d'opérations, il est important de regrouper les données médicales au sein du progiciel et d'utiliser l'outil de sauvegarde intégré au progiciel. Pour limiter les risques de perte de données entre deux sauvegardes, une automatisation du processus de sauvegarde est souhaitable. Les méthodes de chiffrement des données sont mal connues des généralistes. Toutes les données médicales destinées à l'export à un tiers doivent être chiffrées. Elles devraient être chiffrées par défaut par les progiciels. Dans le cas d'un progiciel en cloud, il faudrait réaliser annuellement une exportation des données du cloud sur un support local, en plus des vérifications faites par le fournisseur de service. En cas de difficulté de gestion de son système de sauvegarde local, ou de volonté d'héberger ses données médicales chez un tiers en plus de ses propres sauvegardes, il est conseillé de se diriger vers un éditeur de progiciel agréé Hébergeur de Données de Santé. L'agrément HDS autorise une redistribution des tâches informatiques et libère du temps pour le généraliste. La réalisation spécifique d'archives informatiques n'est ni obligatoire ni nécessaire. Cependant, il est obligatoire d'avoir un système de sauvegarde fiable englobant l'intégralité des données médicales et de respecter les délais de conservation conseillés par l'Ordre des Médecins. Vérifier la réussite du processus de sauvegarde permet de repérer une erreur critique corrigible. Pouvoir paramétrer le logiciel de sauvegarde pour afficher automatiquement un message en cas d'erreur du processus de sauvegarde est important.

MOTS CLÉ

Protection, sauvegarde, archive, données, médicales, informatique, progiciel, médecine générale, support physique, *cloud*, externalisation, restauration, test, chiffrement, risque, responsabilité, confiance, plan de restauration, continuité, compétence, perte de données, règle des 3-2-1, automatisation, résilience, Hébergeur agréé de Données de Santé (HDS), RGPD, Raspberry Pi, Limesurvey, linux, Brahmi, Cheriet, Staechelé, URPS

JURY Président : Monsieur le Professeur LETRILLIART Laurent
Membres : Monsieur le Professeur COLIN Cyrille
Monsieur le Professeur FANTON Laurent
Monsieur le Docteur BLANCHET-MAZUEL Loïc

DATE DE SOUTENANCE 09/05/2019

ADRESSE POSTALE DE L'AUTEUR : 14 C rue du Docteur Rebatel, 69003 LYON

E-MAIL DE L'AUTEUR : drpaul.maguy@laposte.net