

Comment préserver l'Assurance Maladie avec la gouvernance des données critiques ?

11 juin 2026

Yann GOUEDO

Head of Data & AI – Direction Technique

yann.gouedo@thalesgroup.com

+33 6 74 55 44 32

www.thalesgroup.com



Sommaire



Sommaire



La Santé concentre toutes les tensions de la donnée critique



Agir ensemble, protéger chacun

QUI SOMMES-NOUS ? | CARRIÈRES | **PRESSE** | ÉTUDES ET DONNÉES | ACTUALITÉS | PODCASTS

Assurance Maladie > Presse > Communiqués et dossiers de presse
> Données de santé : état des lieux et enjeux des bases de données

Filtres

Votre recherche
mot-clé

Rechercher par catégorie :

- Accès aux droits / Accès aux soins
- Dépenses de santé
- Efficience des soins
- Publications / Études / Données
- Risques professionnels
- Santé et prévention
- Services
- Vie conventionnelle
- Vie de l'institution

Par région :

- Communiqués nationaux
- Communiqués régionaux

Par période :

de MM AAAA
à MM AAAA

Données de santé en France : état des lieux et enjeux des bases de données, rôle de l'Assurance Maladie

20 avril 2023 [X](#) [f](#) [in](#) [< RETOUR](#)

[PUBLICATIONS / ÉTUDES / DONNÉES](#)

Aujourd'hui, les enjeux autour des grandes bases de données de santé, de leurs accès et de leurs exploitations, comme de leurs protections, font l'objet de beaucoup de discussions et d'attentes, voire même de craintes ou d'idées préconçues.

Cela vient du fait qu'il s'agit d'un sujet qui combine hyper technicité et enjeux stratégiques.

Car oui, le sujet est très technique, en raison de l'immensité des bases en jeu - on parle ici de téraoctets et de milliards de lignes. De plus, pour gérer et exploiter ces données, des systèmes d'information et des outils statistiques tout aussi hors normes sont nécessaires.

Mais les fortes attentes autour de ce sujet viennent avant tout des perspectives nouvelles associées à leurs exploitations comme des enjeux stratégiques qu'elles représentent pour notre système de santé. Et de fait, les données de santé et leurs accès sont un formidable vecteur d'innovation en santé. Celles-ci sont exploitables à l'infini dans le champ de la recherche et de l'évaluation des prises en charge et de l'information de tous – décideurs, professionnels de santé mais aussi patients. Tous ces bénéficiaires expliquent l'appétit que ces données aiguïsent.

VOLUME

1,5 milliards de feuille de soins / an (téraoctets de données)

CRITICITE

Dépendances financières, soupçons de fraude, atteinte à la vie privée

SENSIBILITE

Etude d'intérêt général avec strict respect RGPD

EXPERTISE

Coffre-fort de savoir et d'expertise sur l'ensemble des spécialités

SOUVERAINETE

Gouvernance partagée sans dilution de responsabilités

PATRIMOINE

Système de données de santé riche et hétérogène au sein d'un écosystème

L'IA ne révèle pas seulement la fraude

Elle est le révélateur de l'état de la gouvernance des données

Elle révèle

■ Ce que les données disent vraiment des processus métier

■ Ce que les silos organisationnels empêchent de voir

■ Ce que la gouvernance doit être

Sommaire



Contexte CPAM Finistère

Un des enjeux de la CPAM du Finistère est d'**optimiser les activités de contrôles et de détections des montants indûment versés** aux bénéficiaires et professionnels de Santé

- ❖ De **nombreuses applications sont utilisées, pour mener des analyses exploratoires et des investigations manuelles**, ce qui nécessite un **investissement en temps important pour identifier du risque**, et ce quel que soit son niveau de criticité (montant indus et capacité à s'étendre)
- ❖ La démarche actuelle présente le **risque de passer à côté de nombreux cas d'anomalies et par conséquent de fraude**
- ❖ **Le nombre de cas de risques avérés est faible voire inexistant** au regard du large nombre de transactions et des montants élevés de remboursements



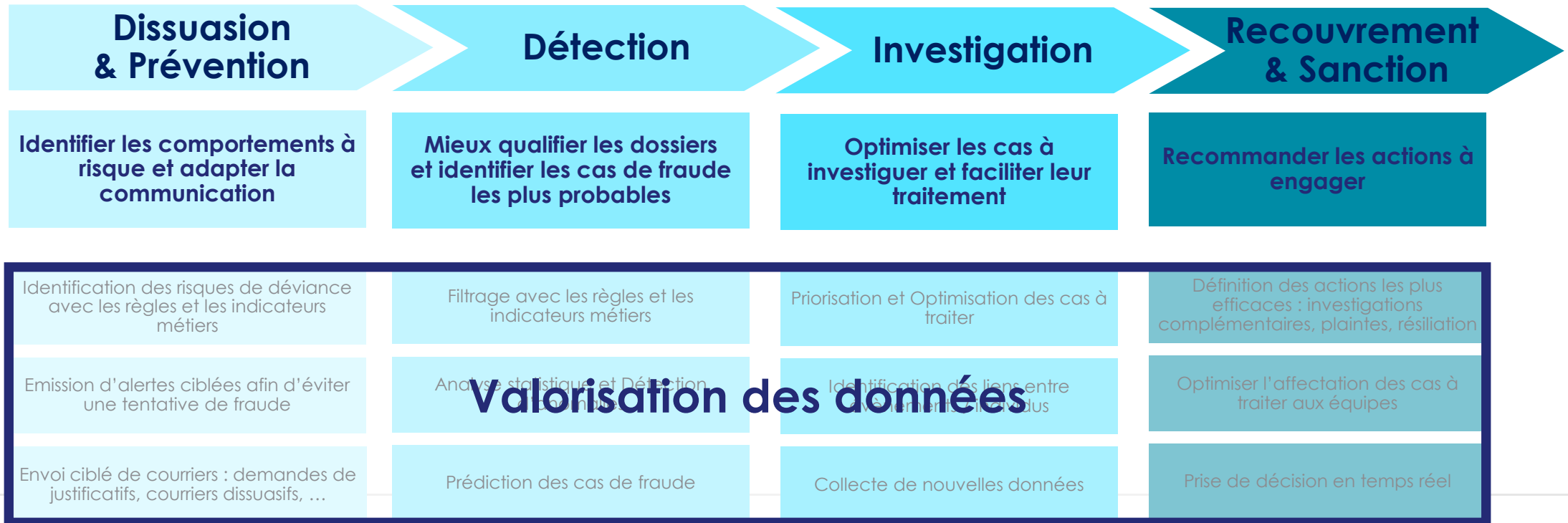
Aligné avec les ambitions de la CNAM, la **CPAM du Finistère souhaite avancer dans sa transformation digitale** avec la valorisation des données par l'Intelligence Artificielle **en réalisant un Proof of Value (PoV) ayant pour objectifs de**

- ❖ **Mettre en évidence des potentiels gains par** une approche de valorisation de données dans la Lutte contre la fraude
- ❖ **Evangéliser les équipes de la CPAM du Finistère** sur l'apport et les leviers de la valorisation des données dans les activités de contrôles et de détection de fraude
- ❖ **Disposer de recommandations** sur les améliorations à mettre en place pour permettre à la CPAM du Finistère d'améliorer les actions de contrôles et de détection

La Valorisation des données avec l'IA offre des leviers d'optimisation à tous les niveaux

Les domaines clés du dispositif de lutte contre la fraude

Processus d'amélioration continue



Périmètre du Proof of Value (PoV)

Les **professions de santé** visées pour la réalisation du PoV sont :



- ❖ Liste produits et prestations (LPP) (pansements; locations; perfusions; ...)



- ❖ Transports médicaux (transports sanitaires; taxis)



Périmètre de données



- ❖ Type de données : données de facturation (Erasmus)
- ❖ Données de facturation : janvier 2023 à juin 2024

Après les premières explorations, la décision de **concentrer les efforts sur la seule profession de santé « Liste produits et prestations (LPP) »** a été prise pour les raisons suivantes :

- ❖ Moyens de détection du risque à renforcer
- ❖ Forte mobilisation d'acteurs CPAM pour avancer sur cette profession de santé
- ❖ Optimisation des échanges et itérations (notamment lors des ateliers bi hebdomadaires), par le focus sur une seule profession

Processus et Données « Liste des produits et prestations (LPP) »

	Prescription	Commande / Achat	Demande de remboursement	Facturation
Périmètre	Produits et Prestations. Dispositifs médicaux pour traitements et matériels d'aide à la vie, aliments diététiques et articles pour pansements, des orthèses et prothèses externes, des dispositifs médicaux implantables et des véhicules pour handicapés physiques.			
Acteurs majeurs	<ul style="list-style-type: none"> Médecins Bénéficiaires 	<ul style="list-style-type: none"> Etablissements Pharmacies Fournisseurs Bénéficiaires 	<ul style="list-style-type: none"> CPAM (opérateurs, contrôleurs) Professionnels Santé Bénéficiaires 	<ul style="list-style-type: none"> CPAM (opérateurs, contrôleurs) Professionnels Santé Bénéficiaires
Points d'attention	<ul style="list-style-type: none"> Véracité des Prescriptions Accessibilité à des pièces justificatives 	<ul style="list-style-type: none"> Analyse du besoin (ordonnance) Droits prise en charge 	<ul style="list-style-type: none"> Contrôles de cohérence Risque 	<ul style="list-style-type: none"> Risque
Données ingérées	Consultations	<ul style="list-style-type: none"> Ordonnances Produits Bénéficiaires Professionnels 	<ul style="list-style-type: none"> Ordonnances Produits Bénéficiaires Professionnels 	<ul style="list-style-type: none"> Décomptes de remboursement Risque
Données générées	Ordonnances	Transactions	Décomptes de remboursements	Facturations

Bénéfices attendus dans la mise en place d'une stratégie d'amélioration de Lutte Contre la Fraude (LCF)

Pour l'Assurance Maladie

- Meilleure compréhension des usages et profils (KYC)
- Réduction de sommes indûment versées par l'augmentation et la meilleure pertinence des contrôles a priori sans dégrader la qualité de service (délai de remboursement)
- Amélioration de l'efficacité des ciblage des contrôles à posteriori
- Amélioration de la productivité des opérations de contrôle et de recouvrement (réduction coûts)
- Raccourcissement du délai nécessaire au contrôle et au recouvrement
- Amélioration de la réactivité face aux évolutions de la fraude
- Meilleure sensibilisation et prévention sur les bonnes pratiques
- Meilleure traçabilité des opérations

Pour les utilisateurs / bénéficiaires & professionnels de santé

- Amélioration de la satisfaction utilisateurs
- Diminution de la répercussion des coûts induits par la fraude
- Gains de productivité
- ...



Ce que l'IA a révélé comme Risque A partir des Données

Résultats PoV et Estimation du Risque

Liste des produits et prestations (LPP)

Identifiés

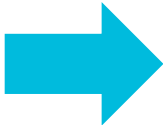
Plus de 400 factures
en anomalies

Pour un montant de 475 571 €

TOTAL	FINISTERE (PRESC + EXEC)
2,3 millions factures	1,8 millions factures
143 M€ (LPP)	86 M€ (LPP)
5 M€ (LPP / Perfusions) pour 49 000 factures	1 M€ (LPP / Perfusions) pour 17 000 factures
521 486 bénéficiaires	475 153 bénéficiaires
39 454 professionnels de santé	7 867 Professionnels de santé

Anomalies confirmées

10,8 % des factures



- **51 735 € Risque confirmé**
(Perfusions 3,5% LPP)
- **Risque estimé LPP = 1,48 M€**
(période de 18 mois)

Période

Janvier 2023 à Juin 2024

Types de risque identifiés

Des remboursements
qui dépassent le
plafond autorisé par
bénéficiaire

Une même écriture
manuscrite sur 2
ordonnances alors que
les prescripteurs sont
différents

Une facturation réalisée
sans respect de la
prescription

Des doubles
facturations d'actes

Des pratiques
suspectes entre
prescripteurs /
bénéficiaires et
prescripteurs /
exécutants

Des abus, des
surconsommations et
fraudes en réseau

Risque financier & Coûts opérationnels

Risque : fraudes, abus, erreurs, ...	Dissuasion & Prévention		Détection		Investigation		Recouvrement & Sanction	
	Existant	PoV	Existant	PoV	Existant	PoV	Existant	PoV
Risque financier (montants en €)								
Suspicion détectées (S)								
Fraudes avérées (F)								
Recouvrements (R)								
Fraudes évitées (E)								
Nouveaux cas de fraudes (N)								
Coûts opérationnels (temps en heures convertis en €)								
Rappel des bonnes pratiques (CO1)								
Gestion des requêtes nationales et locales (CO2)								
Investigations (CO3) - Signalements, Requêtes								
Contrôles fraude (CO4) - Contrôleurs et agents comptables								
Recouvrements (CO5)								
Demandes (pré-étude; analyses ...) (CO6, CO7)								
Coûts réalisation (montants en €)								
Réalisation PoV Thales								

a (élargissement par similarités) = coefficient représentant la possible part de fraude additionnelle provenant des investigations menées

b, c, d (recouvrement) = coefficients pondérant la fraude détectée en vue d'estimer les sommes indues récupérées

e, f (dissuasion) = coefficient pondérant les possibles fraudes évitées grâce aux rappels des bonnes pratiques

Légende

	Existant (CPAM)
	Montants confirmés POV (THALES)

Sommaire





Ce que l'IA a révélé Sur la Gouvernance

Difficultés liées aux données et informations utiles

Constats

- ❖ **Large Hétérogénéité des données**, provenant de sources variées
- ❖ **Qualité des données mitigée**, avec beaucoup de valeurs manquantes
- ❖ **Accès, confidentialité et qualification des acteurs** (avec le respect des contraintes réglementaires RGPD)
- ❖ **Peu ou pas d'historique de cas avérés de fraude**



Perspectives

- ❖ **Architecture Data centrée** avec un entrepôt de données dédié à la détection de la fraude, avec règles de cohérence & contrôles
- ❖ **Connaissance des acteurs (KYC Know Your Customer)** : prescripteurs, exécutants et bénéficiaires (respect RGPD)
- ❖ **Enrichissement des données** de facturations avec indicateurs et open data
- ❖ **Historisation des cas avérés de fraude** en lien avec OG3S

Défis en termes de moyens de valorisation des données

Constats

- ❖ **Requêtes de détection du risque**
 - Requêtes locales et nationales (concernant très peu de codes LPP)
 - Tableur manuel cartographiant le risque (règles non implémentées)
- ❖ **Outil de Profiling des professionnels de santé (Profileur)**
- ❖ **Approche statistique et exploratoire**
- ❖ **Processus manuel de détection** et en support des investigations
- ❖ **Peu ou pas d'historique de cas avérés de fraude**



Perspectives

- ❖ **Renforcement et élargissement des requêtes**
- ❖ **Exploitation des informations de Profiling** au travers d'un entrepôt de données dédié à la détection de fraude
- ❖ **Automatisation de plusieurs moyens de détection non supervisés** (statistiques, détection d'anomalies, graphes ...)
- ❖ **Apprentissage automatique sur les fraudes avérés** (classification)

Enjeux en termes d'organisation

Constats

- ❖ **Nécessité de multiples expertises dédiées à la LCF** (frais de santé, contrôles, données et outils, et manque de clarté sur les responsables des contributions)
- ❖ **Adhésion mitigée** des parties prenantes sur **l'Intelligence Artificielle** pour améliorer la détection de fraude
- ❖ **Absence de compétences en science de données et intelligence artificielle**



Perspectives

- ❖ **Nécessité de structurer une équipe intégrée (métiers et IT)** dédiée à la détection de fraude en alignement avec la feuille de route
- ❖ **Partage** de retours d'expérience et des points d'attention
- ❖ **Acculturation** aux données et à l'intelligence artificielle

Retour sur Investissement (ROI)

Constats

- ❖ **Difficultés à démontrer le ROI**
 - Durée courte pour mener les itérations Détection ↔ Investigation
 - Moyens limités
 - Mobilisation limitée des différentes expertises nécessaires
- ❖ **Cas d'anomalies et de risque ambigus et complexes**
- ❖ **Démarche non généralisable à l'échelle d'une profession**



Perspectives

- ❖ **Nécessité d'allonger la durée du PoV (et donc le budget)** afin de permettre de trouver plus de cas probants, et d'élargir fonctionnellement
- ❖ **Nécessité d'un travail collaboratif** entre experts métiers et data scientists (travail en équipe intégrée)
- ❖ **Nécessité d'affiner le cadrage** pour la bonne poursuite au travers de la feuille de route

Sommaire

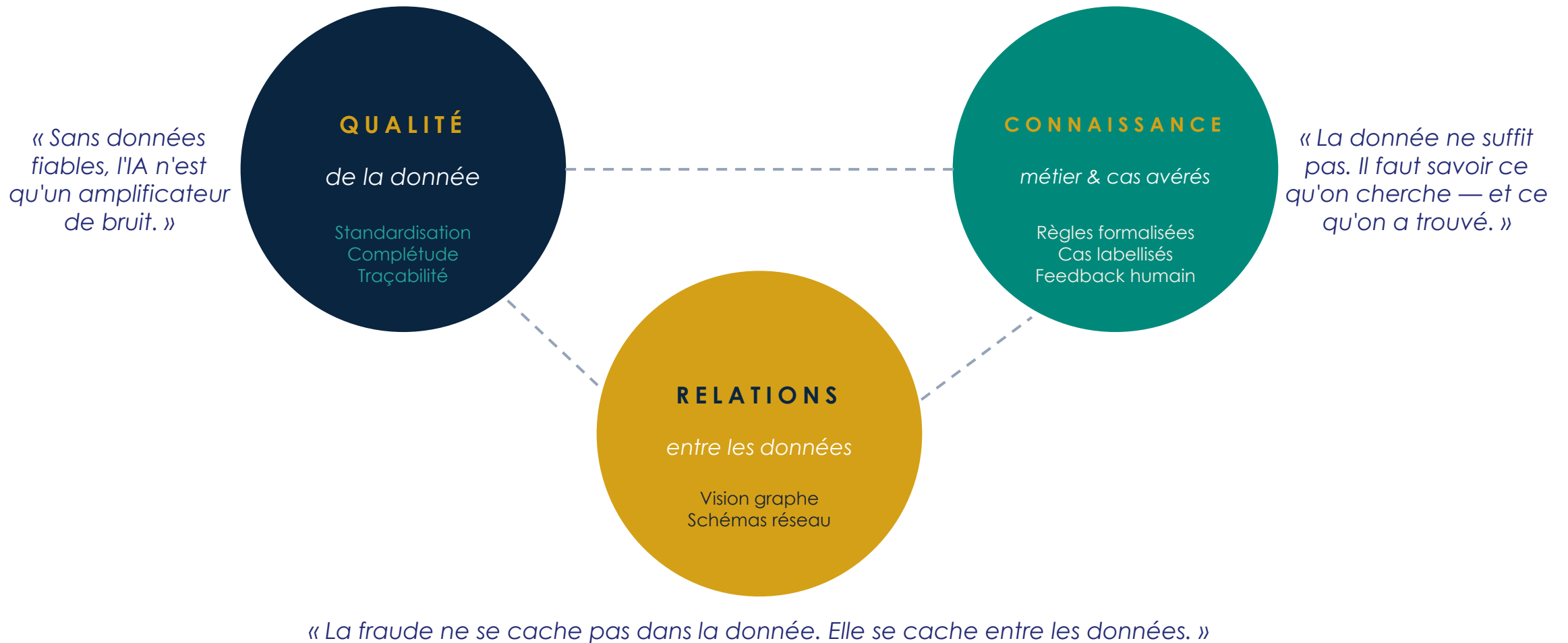




**L'IA n'a pas seulement permis de détecter
du Risque**

**Elle a révélé ce qu'il est nécessaire de
structurer pour passer à l'échelle**

Un triptyque de gouvernance pour l'ère de l'IA



Feuille de route CPAM : chantiers identifiés

Thème	Initiative		Prérequis	Charge estimée
Data Science & AI	1 Identification des professions et périmètres	<ul style="list-style-type: none"> Recenser les flux de données utiles pour adresser la lutte contre la fraude, et les données afférentes Prioriser les professions de santé et les périmètres d'études, selon la capacité à identifier du risque avec la valorisation des données Documenter la feuille de route fonctionnelle et la structure du ROI 	<ul style="list-style-type: none"> Retour d'expérience PoV Acculturation méthodologie Data Thinking et ses objectifs Implication de toutes les expertises 	[Charge estimée]
	2 Architecture data centrée	<ul style="list-style-type: none"> Recenser l'existant SI (technique, fonctionnel, organisationnel) Définir la structure et la gouvernance d'un entrepôt de données dédié à la lutte contre la fraude Rédiger un Dossier d'Architecture Technique (DAT) Mettre en place et alimenter l'entrepôt de données 	<ul style="list-style-type: none"> Recensement existants SI Recensement compétences internes 	
	3 Enrichissement des données	<ul style="list-style-type: none"> Identifier les informations utiles pour enrichir les données originelles Définir et implémenter ces nouvelles informations Définir et implémenter des indicateurs statistiques Réaliser des analyses descriptives de l'ensemble des données et informations 	<ul style="list-style-type: none"> Retour d'expérience PoV Implication de toutes les expertises 	
	4 Exploitation de sources de données ouvertes (open data)	<ul style="list-style-type: none"> Identifier des sources de données ouvertes (open data) Définir la gouvernance de ces sources open data Alimenter l'entrepôt de données avec ces sources de données 	<ul style="list-style-type: none"> Recensement des idées existantes 	
	5 Méthodes et outils de valorisation des données	<ul style="list-style-type: none"> Identifier les critères pour le choix des méthodes et outils Définir les outils nécessaires à la lutte contre la fraude Mettre en place les outils, en connexion avec les outils existants Acculturer les personnes impliquées à la donnée et l'intelligence artificielle 	<ul style="list-style-type: none"> Recensement existant Retour d'expérience PoV Recensement compétences & expertises 	
	6 Expérimentations par périmètre fonctionnel	<ul style="list-style-type: none"> Identifier et définir les acteurs à mobiliser selon le périmètre fonctionnel à adresser Réaliser les expérimentations sous forme d'itérations avec les ateliers 	<ul style="list-style-type: none"> Recensement compétences & expertises Structuration en équipe intégrée 	
	7 Capitalisation avec IA supervisée (historisation des cas avérés)	<ul style="list-style-type: none"> Engager une démarche itérative « anomalies & investigations » avec l'objectif de statuer des cas avérés (fraude, erreur, abus, ...) Alimenter l'entrepôt de données avec les cas avérés Appliquer des modèles IA supervisés pour automatiser la détection du risque 	<ul style="list-style-type: none"> Forte collaboration entre les métiers et les experts techniques 	

THALES

Building a future we can all trust



www.thalesgroup.com