

# Le défi des données pour l'inspection générale des affaires sociales



Juliette Berthe\*

L'inspection générale des affaires sociales (Igas) mène des missions d'audit et d'évaluation dans les domaines de la santé, du travail et des solidarités. Elle s'est dotée d'un pôle *data* pour asseoir les travaux de l'inspection sur des éléments quantifiés et objectiver en particulier certains points ou recommandations. Les *data scientists* doivent exploiter des données sur des thématiques variées, répondre aux besoins des missions dans des délais courts et composer à la fois avec des bases bien structurées, notamment dans le champ de la santé et de l'emploi, et des systèmes d'information locaux hétérogènes et fragmentés, notamment dans le domaine des solidarités.

Le travail du *data scientist* est vaste : il doit rechercher, croiser et nettoyer des données qui peuvent être incomplètes ou dispersées, en garantissant leur pertinence et leur interprétation correcte. Lorsqu'elles sont absentes, il recourt à des approches comme le *web scraping*. Il doit aussi arbitrer entre fraîcheur et fiabilité des données, privilégiant parfois des sources officielles mais datées pour assurer la robustesse des analyses.

Son rôle ne se limite pas à la *data science* : il interagit avec les inspecteurs et les acteurs de terrain pour contextualiser ses analyses. En combinant rigueur, adaptation et pragmatisme, il éclaire les décisions publiques malgré des contraintes méthodologiques et temporelles fortes, dans l'objectif de contribuer à des politiques plus informées et efficaces.

 The General Inspectorate of Social Affairs (IGAS) conducts audit and evaluation missions in the fields of health, labor, and social solidarity. It has set up a data pole to consolidate the inspectorate's projects by using quantified elements and particularly to substantiate some points or recommendations. Data scientists have to process thematically diverse data, meet the mission needs within short timeframes, and work with well-structured databases, in the fields of health and employment notably, as well as with heterogeneous and fragmented local information systems, as is particularly the case with social solidarity.

The data scientist has a wide scope of activities: they have to search for, cross-reference, and clean data that is sometimes incomplete or scattered, ensuring its relevance and accurate interpretation. When data is missing, they turn to alternative approaches such as web scraping. They also have to balance between data freshness and reliability, sometimes favouring official but older sources to ensure the robustness of analyses.

Their role extends beyond data science: they collaborate with inspectors and field actors to provide context to their analyses. By combining rigor, adaptability, and pragmatism, they help inform public decision-making despite strong methodological and time constraints, ultimately aiming to contribute to more effective and well-informed policies.

\* Responsable du pôle *data*, inspection générale des affaires sociales (Igas).  
[juliette.berthe@igas.gouv.fr](mailto:juliette.berthe@igas.gouv.fr)

Pour orienter leurs choix stratégiques, les décideurs publics ont besoin d'analyses sérieuses et rigoureuses. Que ce soit pour concevoir des politiques publiques, estimer l'impact d'une mesure, élaborer de nouvelles lois ou allouer des ressources, ils doivent s'appuyer sur des études approfondies et impartiales menées par des spécialistes du sujet. Les inspections ministérielles apportent cette expertise précieuse, chacune dans son domaine spécifique, ou en collaboration lorsque la complexité du sujet exige une approche multidisciplinaire. Dans le domaine social, cette responsabilité revient à l'**inspection générale des affaires sociales<sup>1</sup>** (Igas). Son périmètre d'intervention couvre des enjeux majeurs mobilisant une part significative des ressources nationales et affectant directement la vie de tous les citoyens : emploi, travail et formation professionnelle, santé publique, organisation des soins, cohésion sociale, sécurité sociale, protection des populations.

La disponibilité croissante de larges bases de données administratives ou d'enquêtes (d'accès public ou sur demande motivée), conjuguée à un contexte d'innovation en matière de méthodes quantitatives, constitue une véritable opportunité pour l'enrichissement et la pertinence des constats et recommandations des missions confiées à l'Igas. L'inspection a donc créé en 2023 un pôle *data* où ces aspects d'analyse sont pris en charge par des *data scientists*<sup>2</sup> dédiés. Il comprend aujourd'hui trois *data*

*scientists* permanents et un étudiant en apprentissage. Il a déjà appuyé vingt missions réparties équitablement sur l'ensemble des champs que recouvre l'inspection.

**Si ces experts de la science des données doivent développer des indicateurs pertinents, dans un esprit de rigueur, de transparence et d'objectivité, leur approche diffère sensiblement des pratiques de la statistique publique sur de nombreux aspects.**

Si ces experts de la science des données doivent développer des indicateurs pertinents, dans un esprit de rigueur, de transparence et d'objectivité, leur approche diffère sensiblement des pratiques de la statistique publique sur de nombreux aspects. En effet, plusieurs spécificités encadrent strictement le travail du *data scientist* : l'ampleur du champ d'action de l'Igas, la nature et la source des données disponibles et, surtout, l'objectif des missions, qui est d'éclairer le décideur,

souvent dans des délais très courts, sur des problématiques généralement très ciblées. Cet objectif impose des contraintes temporelles et une utilisation précise des données. Pourtant, c'est justement ce cadre exigeant, à la fois stimulant et complexe, qui constitue la spécificité et l'intérêt de ce travail et son caractère spécial.

## ► Une culture de la donnée très variable selon les champs —

L'Igas opère dans trois domaines : le travail, la santé et les solidarités. La richesse des ressources en données et leur niveau de maturité varient grandement selon ces champs. Les données portant sur le travail et la santé sont les plus robustes.

<sup>1</sup> <https://igas.gouv.fr/>.

<sup>2</sup> Voir la définition du métier de *data scientist* dans Bourlange et al. (2021).

## La santé, très acculturée aux données

Dans le secteur de la santé, la culture de la donnée est ancrée depuis longtemps chez la plupart des acteurs. À l'hôpital, la saisie de données à des fins à la fois financières et épidémiologiques a débuté dans les années quatre-vingts. Au fur et à mesure, ces saisies ont porté sur davantage d'informations ou ont été enrichies d'autres données. Aujourd'hui, l'intérêt de faire remonter de l'information et de la consolider est partagé par l'ensemble des acteurs et le processus fait partie de leur quotidien. Par ailleurs, les données administratives sont particulièrement riches grâce au **système national des données de santé<sup>3</sup>** (SNDS) (*encadré 1*). Ce dernier contient notamment des informations très détaillées sur la consommation de soins à la maille du patient, que ce soit des soins en ville ou en établissement de santé. Il s'agit principalement du détail des remboursements des dépenses par l'assurance maladie et des données décrivant l'ensemble des séjours à l'hôpital. Par ricochet, ce niveau de détail renseigne très précisément sur l'activité des professionnels et des établissements. Les variables financières des établissements publics sont également communiquées et consolidées à une échelle nationale.

*Les données contenues dans le SNDS sont structurées et standardisées. Elles obéissent à des nomenclatures établies et partagées nationalement, voire internationalement.*

Les données contenues dans le SNDS sont structurées et standardisées. Elles obéissent à des nomenclatures établies et partagées nationalement, voire internationalement pour certaines. Par exemple, les causes de décès sont codées selon la classification internationale des maladies de l'Organisation mondiale de la santé (Coudin et Robert, 2024). Grâce à cette standardisation, l'interopérabilité interne et externe du système est garantie.

En complément de ces données individuelles, les agences régionales de santé (ARS) utilisent divers outils de gestion pour piloter l'offre de soins, gérer les financements et suivre les dépenses, notamment ceux des établissements. Enfin, les enquêtes et études menées en particulier par la direction de la recherche, des études, de l'évaluation et des statistiques (Drees) apportent des données particulièrement précieuses sur une multitude de sujets qui vont au-delà du champ de la santé à strictement parler ; le champ médico-social y est notamment abordé. Une grande part de ces données est disponible en open data sur son site. Les données plus confidentielles, notamment celles relevant d'enquêtes spécifiques, sont pour la plupart disponibles au Centre d'accès sécurisé aux données (CASD) [Gadouche, 2019].

## Le champ travail possède également des données riches et structurées

La culture de la donnée est également très présente dans le domaine du travail au sens large ; on parle alors du champ TEFP (travail, emploi et formation professionnelle). La principale source de données administratives est la **déclaration sociale nominative** (DSN) : elle apporte des informations individuelles sur les salariés, leurs contrats de travail, les rémunérations et primes reçues, les cotisations versées, les absences et

<sup>3</sup> <https://www.snds.gouv.fr/SNDS/Accueil>.

## ► Encadré 1. Le SNDS retrace le parcours de soins de chaque assuré social

En France, le système national des données de santé (SNDS) rassemble les principales données de santé provenant de sources administratives et médicales. Géré par la plateforme des données de santé (PDS ou *Health data hub*)<sup>\*</sup> et la caisse nationale de l'assurance maladie (Cnam), il est issu de la mise en relation du système national d'information interrégimes de l'assurance maladie (Sniiram), du programme de médicalisation des systèmes d'information (PMSI) et de la base des causes médicales de décès (BCMD).

Le Sniiram, géré par la Cnam, contient les données relatives à toutes les dépenses d'assurance maladie. Il est composé d'une base de données individuelles sur la consommation de soins, appelée *datamart*<sup>\*\*</sup> de consommation interrégimes (DCIR), de 15 bases thématiques de données agrégées et d'un échantillon au 2/100<sup>e</sup> de patients ayant bénéficié d'un soin, destiné à des études longitudinales.

Le PMSI, géré par l'Agence technique de l'information sur l'hospitalisation (ATIH), a été créé en 1982 pour permettre l'analyse de l'activité médicale des établissements de santé à des fins d'allocation budgétaire. Il intègre des informations administratives et médicales relatives à chaque séjour dans un établissement de santé, public ou

privé, pour tout type d'hospitalisation : médecine, chirurgie et obstétrique (MCO), soins médicaux et de réadaptation (SMR), hospitalisation à domicile (HAD) et psychiatrie.

Enfin, la BCMD, gérée par le Centre d'épidémiologie sur les causes médicales de décès (CépiDC), contient les données relatives aux causes de décès pour chaque individu.

En résumé, le SNDS fournit des informations détaillées :

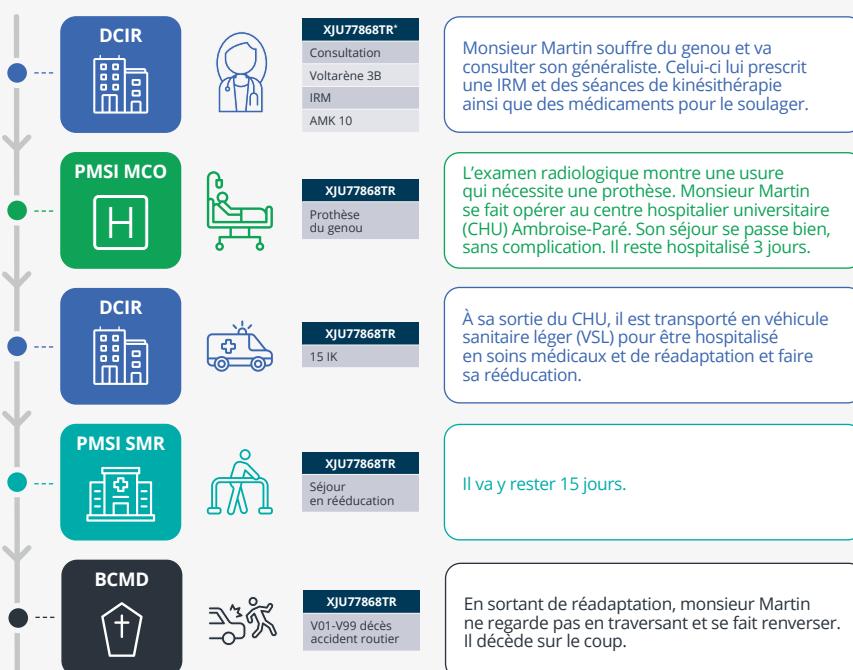
- sur l'ensemble des dépenses liées aux soins, qu'ils soient réalisés en ville ou en établissement de santé et quel que soit le type de dépense : consultation médicale ou paramédicale, dépense de pharmacie, transport sanitaire, etc. ;
- sur l'ensemble des séjours en établissement : durée, passage aux urgences, coût, mais aussi diagnostics et comorbidités du patient.

Les trois systèmes Sniiram, PMSI et BCMD permettent, par le biais d'un identifiant unique, de chaîner les données à la granularité du patient. On obtient ainsi une vision globale du parcours de soins des assurés sociaux jusqu'à la cause de leur décès (**figure encadré**).

\* <https://www.health-data-hub.fr>.

\*\* Magasin de données.

### Le parcours de soins de monsieur Martin dans le SNDS



\* XJU77868TR est l'identifiant (fictif) de monsieur Martin dans le SNDS. C'est cet identifiant unique dans l'ensemble des bases du SNDS qui permet de suivre son parcours de soin.

reprises, etc. (Humbert-Bottin, 2018). Très normalisée, mais nécessitant beaucoup de retraitements pour la production de statistiques, elle est exploitée à cette fin par une multitude d'acteurs, notamment par l'Insee, qui produit à partir de celle-ci la **base Tous salariés**<sup>4</sup> (Brunet et al., 2023). Si l'Igas n'exploite pas encore directement la DSN, elle s'appuie très souvent sur la base Tous salariés.

Ces données sur l'emploi des salariés issues des processus administratifs sont complétées par d'autres données de gestion sur l'emploi non salarié, d'une part, et sur le chômage et la formation, d'autre part, provenant notamment de la direction générale à l'emploi et à la formation professionnelle (DGEFP), de France Travail et de France compétences. Concernant la formation professionnelle, la Caisse des dépôts a par ailleurs développé une plateforme de mutualisation et d'échange de données, **AGORA**<sup>5</sup>. Son objectif est de consolider en temps réel les informations sur les parcours et d'améliorer ainsi le pilotage dans le domaine. Elle est alimentée par l'ensemble des acteurs impliqués : organismes formateurs, financeurs, rémunérateurs et certificateurs. Enfin, des études et enquêtes menées notamment par l'Insee et la direction de l'animation de la recherche, des études et des statistiques (Dares) complètent la vision du champ. La plupart de ces données sur le champ TEFP sont accessibles par l'Igas, soit en open data, soit par le CASD après un accord des producteurs.

## Les solidarités : un champ plus diversifié et moins centralisé

Le champ des solidarités est particulièrement diversifié ; il regroupe des acteurs et des activités très variés. De manière générale, il concerne le soutien aux personnes vulnérables en raison de leur âge, de leur handicap ou de leur situation familiale ou sociale complexe. Les politiques concernées sont notamment celles de l'accueil de la petite enfance, de la protection maternelle et infantile, de la protection de l'enfance, du soutien aux personnes handicapées ou en perte d'autonomie ainsi qu'à leurs proches aidants, de la lutte contre la pauvreté, de l'aide aux familles nombreuses et aux familles monoparentales. Les interventions relèvent à la fois du soutien financier, de l'accompagnement, de l'aide à domicile, de l'accueil en institution, voire de la prévention et de l'éducation.

*Sur le volet médico-social, la mise à disposition et l'exploitation des données sont plus difficiles.*

La diversité des publics et des actions menées conduit à mobiliser un large éventail d'acteurs, à tous les niveaux de l'échelle territoriale. Au niveau local, les communes et les départements sont chefs de file de l'action sociale : ils jouent un rôle clé dans le financement et la gestion des services sociaux et médico-sociaux. En

région, les ARS se concentrent sur la qualité et la régulation de l'offre médico-sociale liée à la santé, tandis que les directions régionales de l'économie, de l'emploi, du travail et des solidarités (Dreets) œuvrent sur la dimension sociale et professionnelle : elles coordonnent les politiques sociales et travaillent notamment sur l'inclusion sociale et professionnelle. Enfin, au niveau national, la Caisse nationale de solidarité pour

<sup>4</sup> <https://www.insee.fr/fr/metadonnees/source/serie/s1998>.

<sup>5</sup> <https://travail-emploi.gouv.fr/agora-la-plateforme-dechange-de-donnees-de-la-formation-professionnelle>.

l'autonomie (CNSA) coordonne les politiques publiques de soutien à l'autonomie sur les aspects de financement, pilotage des maisons départementales des personnes handicapées (MDPH)<sup>6</sup>, planification et suivi des politiques médico-sociales. La caisse nationale des allocations familiale (Cnaf) et la mutualité sociale agricole (MSA) gèrent quant à elles les prestations sociales à travers leurs réseaux de caisses.

Pour étudier le champ des solidarités, l'Igas a à sa disposition les enquêtes et publications de la Drees et les données de gestion des collectivités et des caisses. On comprend aisément que l'organisation complexe du champ, à la fois par la multiplicité des domaines d'intervention et par l'autonomie plus ou moins grande laissée aux acteurs locaux dans la gestion, conditionne fortement la disponibilité des données, leur homogénéité et leur interopérabilité. D'une manière générale, on observe que l'existence d'un réseau dédié piloté permet des remontées harmonisées. Ainsi, les prestations familiales et de solidarité qui sont versées par les réseaux de la Cnaf et de la MSA sont agrégées dans une base nationale à laquelle l'Igas a accès via le CASD. En revanche, sur le volet médico-social, la mise à disposition et l'exploitation des données sont plus difficiles. En effet, les canaux pour saisir et remonter l'information sont multiples et ne sont pas tous harmonisés. La diversité des acteurs (MDPH, ARS, conseils départementaux, centres communaux d'action sociale) et le fait qu'ils ne soient pas entièrement dédiés à la problématique médico-sociale complexifient le sujet. Sur le champ du handicap, la CNSA travaille à moderniser la gestion des données médico-sociales, notamment par la création d'un système d'information harmonisé et la mise en place d'un centre de données. Elle vise en particulier à construire un système d'information sur les MDPH, permettant d'homogénéiser les informations autour de l'accompagnement par ces services des personnes handicapées, afin d'en donner une vision nationale. Cette base de données contiendra les informations sur les différentes demandes d'accompagnement formulées pour les personnes handicapées et le déroulement de leur instruction. Un appariement avec le SNDS est par ailleurs prévu pour pouvoir observer les liens entre ces accompagnements et le parcours de santé des bénéficiaires.

Plus largement, même si la Drees s'attache à centraliser des informations locales via des enquêtes ou des remontées de données individuelles<sup>7</sup>, il subsiste que les données relatives au champ des solidarités ne bénéficient pas d'une normalisation comme celles de la santé ou du travail. Leur analyse est donc bien plus complexe (Cotton et Haag, 2023) et la robustesse des résultats moins assurée.

<sup>6</sup> Les MDPH sont, dans chaque département, le guichet unique d'accès simplifié aux droits et prestations pour les personnes handicapées.

<sup>7</sup> Par exemple via l'enquête Aide sociale réalisée auprès des départements (Diallo et al., 2024) ou les remontées de données individuelles sur l'orientation, l'accompagnement et l'insertion des bénéficiaires du revenu de solidarité active (RSA).

## ► Les spécificités du travail des *data scientists* à l'Igas

Intervenir dans un contexte de missions, sur des sujets très divers, des problématiques très précises et dans une fenêtre de temps limitée, conditionne fortement le travail des *data scientists* de l'Igas.

### Des délais contraints qui obligent aux compromis

Le rôle du pôle *data* est d'apporter un éclairage spécifique à certaines missions de l'Igas au travers d'analyses quantitatives. Il peut s'agir d'une statistique qui permettra d'appuyer une recommandation, de détection d'anomalies pour mieux cibler des contrôles, d'une classification pour dégager des comportements ou situations similaires, etc. Chaque approche *data* répond à un besoin précis pour la mission.

Les données sur un sujet sont parfois riches et abondantes. Pour autant elles ne répondent pas nécessairement aux nécessités des missions de l'Igas. Leur fraîcheur est notamment un problème récurrent. L'Igas s'appuie beaucoup sur des données officielles, issues d'enquêtes ou de traitements de données administratives réalisés par la statistique publique (Insee ou services statistiques ministériels, SSM). La qualité des données est alors indiscutable, mais la contrepartie réside dans le délai d'obtention. Le risque est que l'Igas dresse des constats ou appuie ses recommandations sur des informations qui

pourraient ne plus refléter, ou que partiellement, la réalité du moment. Il est donc nécessaire que le *data scientist* apprécie avec les inspecteurs de la mission les enjeux de temporalité.

**“ Les données sur un sujet sont parfois riches et abondantes. Pour autant elles ne répondent pas nécessairement aux nécessités des missions de l'Igas.**

Ainsi, dans le cadre d'une mission réalisée en 2024 visant à évaluer le caractère contraint des temps partiels dans certains secteurs d'activité, l'Igas s'est appuyée sur la base Tous salariés de l'Insee. L'objectif était de mesurer la part des salariés travaillant à temps partiel dans certains secteurs choisis, mais aussi le nombre d'emplois cumulés par ailleurs par chacun de ces salariés, pour mieux appréhender le caractère contraint de leur temps partiel (Magnier et Viossat, 2024). À l'époque où les travaux ont été réalisés, le millésime le plus récent de la base Tous salariés était 2022, pour un rapport publié fin 2024 : les constats portaient donc sur des données datant de deux ans. Il aurait été possible de travailler sur des informations plus actuelles issues directement de la DSN. Cependant, les analyses auraient alors reposé sur des données administratives brutes, donc de qualité moindre, notamment sur le nombre d'heures travaillées ; ceci aurait alors fragilisé les constats. L'Igas a finalement choisi de privilégier la fiabilité de l'information, quitte à s'appuyer sur des données moins récentes.

## Savoir trouver les données disponibles et les mettre en musique

Une partie importante du travail du *data scientist* consiste à explorer la multitude de bases de données disponibles, qu'il s'agisse de ressources accessibles en open data ou via des dispositifs sécurisés comme le CASD. Il est en effet crucial de trouver la bonne source d'information. La mission « Lieux de vie et accompagnement des personnes âgées en perte d'autonomie » (Emmanuelli et al., 2023) est un exemple parlant sur le potentiel d'analyse que représentent les données en accès libre. En effet, en mobilisant les informations diffusées sur les sites de la Drees et de l'Insee, il a été possible de caractériser à l'échelle départementale la demande et l'offre de prise en charge de personnes dépendantes :

- d'un côté, les données en accès libre ont permis de décrire comment les personnes âgées dépendantes se répartissent dans les départements, quelle part elles représentent dans la population totale de chaque département et comment la situation est susceptible d'évoluer à l'horizon 2030 ou 2040 ;
- d'un autre côté, les départements ont pu être classés selon quatre catégories d'offre de prise en charge, à partir de sept indicateurs sur l'offre en ville et en établissement.

La confrontation de ces deux analyses permet ainsi à chaque département de se situer du point de vue de l'offre et de la demande (*figure*) : elle constitue un outil concret pour se préparer au défi démographique qui arrive. Par ailleurs, comme ce travail s'appuie sur des données accessibles à tous, les acteurs locaux sont en capacité de le reproduire et de l'actualiser facilement. Ils peuvent également le compléter à partir d'informations qui leur sont propres ou qu'ils jugeraient pertinentes.

## Savoir exploiter les données brutes locales : en investissant dans la connaissance du terrain...

S'il est confortable de travailler sur des données structurées et bien normées, comme celles évoquées précédemment, elles ne suffisent pas toujours à répondre aux besoins diversifiés de l'Igas. Les données des systèmes d'information (SI) locaux sont alors une

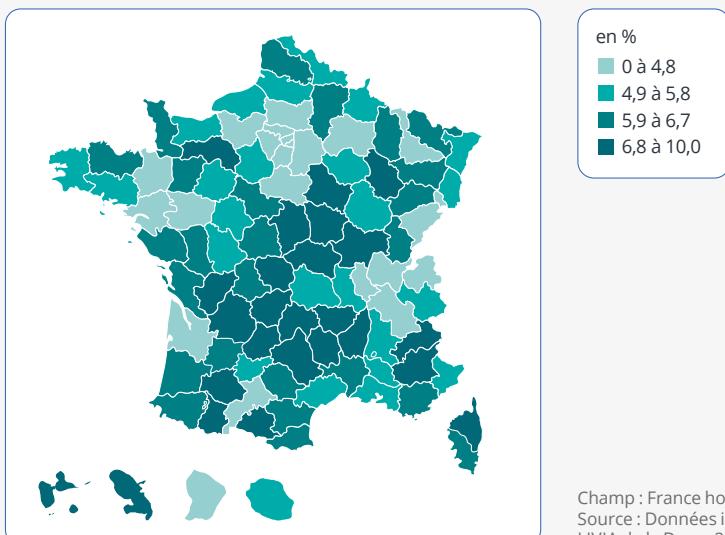
source précieuse d'information. Cependant, elles nécessitent un véritable investissement en vue de leur utilisation. En effet, il s'agit de données brutes de production qui n'ont pas été élaborées à des fins statistiques. Le *data scientist* doit donc réaliser un important travail pour en comprendre le sens, les adapter sur le plan sémantique et réaliser des ajustements techniques, ceci afin d'homogénéiser la structure des fichiers et des variables. Par ailleurs, les SI étant alimentés en permanence, il est essentiel de définir la temporalité sur laquelle va porter l'analyse et de figer les données à un instant  $t$ . Les situations individuelles qui y figurent ne sont alors pas nécessairement à jour : il faut tenir compte de cet aspect dans l'analyse.

***S'il est confortable de travailler sur des données structurées et bien normées, elles ne suffisent pas toujours à répondre aux besoins diversifiés de l'Igas. Les données des systèmes d'information (SI) locaux sont alors une source précieuse d'information qui nécessite un véritable investissement en vue de leur utilisation.***

## ► Figure - Le potentiel d'analyse des données en accès libre : illustration

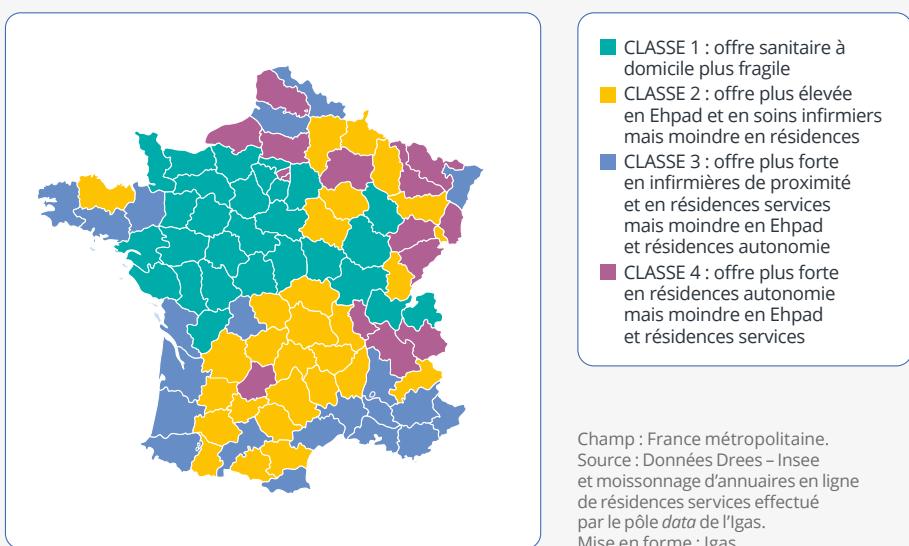
Dans le cadre de la mission « Lieux de vie et accompagnement des personnes âgées en perte d'autonomie » (Emmanuelli et al., 2023), il était important d'éclairer la situation des départements en matière d'offre et de demande de prise en charge de personnes âgées dépendantes. Les données en accès libre de la Drees et de l'Insee permettent d'éclairer les besoins à venir et les forces et faiblesses de l'offre actuelle.

### a Part des personnes âgées en perte d'autonomie dans la population en 2040



### b Caractéristiques de l'offre de prise en charge des personnes âgées en perte d'autonomie

Selon sept variables caractéristiques de l'offre de prise en charge des personnes âgées en perte d'autonomie, une typologie en quatre classes a été construite à partir d'une méthode statistique de regroupement de départements proches.



Cette approche est relativement fréquente, notamment dans le champ médico-social où les données nationales ne sont pas toutes consolidées. Ainsi, dans les missions portant sur l'aide sociale à l'enfance (ASE), l'Igas n'a, pour certains départements, d'autre source que celle des SI des conseils départementaux<sup>8</sup> (**encadré 2**). L'investissement du *data scientist* pour comprendre les données contenues dans ces SI est essentiel : il passe par

## ► Encadré 2. La difficulté à exploiter les données extraites des SI locaux : l'exemple de l'aide sociale à l'enfance (ASE)

L'exemple du suivi par les départements de l'aide sociale à l'enfance (ASE) illustre les difficultés que peut rencontrer l'Igas pour exploiter à des fins statistiques les données des systèmes d'informations locaux.

### Les départements n'utilisent pas tous les mêmes logiciels de gestion

Pour le suivi des jeunes de l'ASE, les départements ont une autonomie totale pour choisir leur logiciel de gestion, qui peut donc varier d'un département à l'autre. Ainsi, dans le cadre de sa dernière mission sur l'ASE, l'Igas a étudié la situation de quatre départements. Trois d'entre eux utilisent le logiciel IODAS, tandis qu'un autre a développé son propre outil, webASE. La précédente mission avait été réalisée dans un département qui utilisait SOLIS. Du fait de cette hétérogénéité, les modalités d'extraction des données, la nature des éléments recueillis et le format des données diffèrent d'un département à l'autre.

### Les logiciels utilisés ne sont pas conçus pour l'extraction massive de données

Les logiciels utilisés pour le suivi des jeunes de l'ASE ne sont pas spécifiquement conçus pour l'extraction et le traitement de grandes quantités de données. Par exemple, IODAS et SOLIS reposent sur l'outil BusinessObject pour exécuter des requêtes, ce qui limite la fluidité et l'efficacité des extractions massives de données. Cette contrainte technique et l'hétérogénéité des logiciels se traduisent par une grande variabilité dans les formats et la structure des informations extraites, avec l'impossibilité d'automatiser d'un département à l'autre les processus d'extraction.

### Les départements n'ont pas tous les mêmes pratiques de gestion des données

Les pratiques de gestion des données varient également selon les départements. Certains d'entre eux adoptent une approche « multidomaine » : un même outil, comme IODAS, est utilisé pour gérer non seulement les données relatives à l'ASE, mais aussi celles concernant d'autres secteurs tels que l'insertion sociale. Cette gestion multiforme des données complique considérablement les demandes d'extraction, qui doivent tenir compte de la diversité des services intégrés dans un même logiciel.

### La qualité des données est également hétérogène et dépend des pratiques

Une fois les données extraites, leur qualité dépend fortement des pratiques de saisie humaine. Selon les préférences des agents en charge du suivi, de nombreux champs sont soit laissés libres, soit remplis selon des règles hétérogènes. Ainsi, des variations typographiques peuvent apparaître pour des informations identiques (par exemple « M. », « Mr » ou « Monsieur » pour une même civilité), ou des incohérences dans le format des dates (par exemple « 01/09/24 » ou « 2024-09-01 »), ce qui complique la standardisation et l'analyse. De plus, certaines erreurs de saisie, ou des conventions locales comme l'utilisation de la date « 2000-01-01 » pour signifier une valeur manquante, rendent difficile la distinction entre des données valides et des données erronées.

### Des données essentielles peuvent manquer (ou être codées en « ne sait pas »)

Un autre problème majeur est l'absence fréquente d'éléments essentiels, concernant notamment les signalements reçus au sujet de l'enfant. Ces données ne sont pas systématiquement renseignées, ce qui empêche une évaluation complète et précise du vécu de l'enfant dans le système de protection. De même, la catégorie des « types de mesure » est souvent incomplète, avec de nombreuses valeurs codées en « ne sait pas », ce qui ajoute à la difficulté d'analyse.

### Les changements de logiciel peuvent rendre complexe l'analyse des évolutions

Enfin, des changements de logiciels, comme celui intervenu en juin 2018 pour un département, peuvent introduire des différences significatives dans la structure des données entre les périodes précédent et suivant la migration. Dans l'exemple évoqué, certaines décisions n'ont pas été transférées lors de la migration : des appariements complexes ont dû être mis en œuvre entre les données anciennes et nouvelles pour assurer leur cohérence. L'introduction du dispositif Olinpe a par ailleurs ajouté de nouveaux champs, qui ne sont pas compatibles avec le format initial du logiciel, ce qui complique encore la gestion des données et leur analyse comparative.

<sup>8</sup> La Drees constitue une base nationale consolidée sur les enfants de l'ASE dénommée Olinpe (Observation longitudinale, individuelle et nationale en protection de l'enfance). Ce dispositif n'est pas encore exhaustif : en 2023, 32 départements ont transmis leurs données.

des échanges constants avec les spécialistes du métier. Il peut également accompagner la mission dans ses déplacements sur le terrain, afin de mieux cerner ce que représentent les données, mais aussi le contexte dans lequel elles sont produites. Cet investissement permet une interprétation éclairée et concrète des phénomènes décrits et une vision critique sur la qualité des données analysées.

### **... et en mobilisant des études nationales, pour une mise en perspective**

La correction manuelle et systématique des données erronées n'est évidemment pas possible ; un choix doit alors s'opérer pour trouver le meilleur compromis entre une qualité d'information acceptable, un nettoyage le moins chronophage possible et une analyse pertinente pour la mission. On peut dans ce cas faire appel à des études nationales, souvent produites par les SSM, pour mettre en perspective dans un contexte plus global les résultats obtenus localement.

Par exemple, dans le cadre des missions portant sur l'ASE, l'Igas s'appuie sur les analyses et l'expertise de la Drees. C'est grâce à ces statistiques qu'elle a pu détecter l'incomplétude des données disponibles dans le SI d'un département. En effet, l'augmentation du nombre d'enfants suivis par l'ASE retracé par le SI était inférieure à celle publiée par la Drees pour ce département, remettant ainsi en cause de manière plus générale l'exhaustivité des données dans les SI. Cette découverte a permis d'alerter l'ASE sur la non-exhaustivité des données dans ses systèmes, la conduisant à s'interroger sur ses procédures de saisie. L'Igas a quant à elle rectifié son approche pour prendre en compte cet élément dans ses analyses.

### **Analyser des pratiques locales pour les étendre à l'échelle nationale**

*Au-delà des résultats d'évaluation à proprement parler, l'un des rôles de l'Igas consiste à réfléchir sur la possibilité pour les acteurs du terrain de mener leur propre analyse.*

Malgré les difficultés évoquées pour les exploiter, les données des SI locaux restent une source incontournable, et ce d'autant plus lorsqu'aucune information n'est disponible au niveau national. Les résultats sur une seule région ou un seul département ne sont certes pas aussi robustes qu'une évaluation faite sur l'ensemble du territoire, mais ils permettent d'émettre une hypothèse de tendance, de montrer la faisabilité d'une analyse, voire d'appuyer une recommandation. En effet, au-delà des résultats

d'évaluation à proprement parler, l'un des rôles de l'Igas consiste à réfléchir sur la possibilité pour les acteurs du terrain de mener leur propre analyse. Plusieurs méthodes peuvent être testées et discutées, l'objectif étant de souligner la faisabilité de l'étude et l'avantage qu'auraient à tirer les acteurs du système à la réaliser sur des données plus complètes ou plus robustes.

La contribution des *data scientists* sur la mission « Les parcours des usagers de la sécurité sociale » (Fournier et al., 2025) illustre bien cette approche. Il s'agissait notamment d'étudier, à la suite de la mise en place d'une offre de service par l'Urssaf<sup>9</sup>, si les contacts que les créateurs d'entreprise avaient avec un agent de l'organisme se traduisaient par un changement de comportement : amélioration de la qualité de leur déclaration et du paiement de leurs cotisations et/ou meilleur exercice de leur droit à une aide sociale. La Caisse nationale des Urssaf a transmis à l'Igas l'ensemble des informations sur les données financières et les contacts entrants (émanant des créateurs d'entreprise), mais elle n'avait pas la trace des contacts sortants (émanant des agents de l'Urssaf). Ces données permettaient donc d'avoir une vision exhaustive sur l'impact des contacts entrants, mais il subsistait une certaine frustration à ne pas pouvoir mesurer l'impact des contacts sortants. Cependant, le SI local de l'Urssaf en région Languedoc-Roussillon disposait de cette information, ce qui a permis de réaliser l'étude dans cette région. Les résultats sont certes limités à ce seul territoire, mais ils montrent la faisabilité de l'analyse et offrent une première estimation pour quantifier l'impact de ces appels. Ils démontrent aussi l'intérêt de tracer cette information, et donc d'étendre cette pratique à l'ensemble des Urssaf. C'est en effet un véritable enjeu pour les caisses régionales, qui pourraient ainsi mieux cibler les destinataires des appels émis par leurs conseillers.

### Savoir aller chercher la donnée quand elle n'est pas recensée ou disponible...

Les SI locaux sont une véritable mine d'informations que l'Igas ne néglige pas, malgré toutes les difficultés inhérentes à leurs exploitations. Toutefois, certaines missions peuvent nécessiter de s'appuyer sur des données qui ne sont consolidées nulle part, en raison de la nouveauté ou du manque d'exploitation du sujet. L'Igas peut alors créer ses propres bases pour répondre au besoin spécifique des missions.

*L'Igas peut créer ses propres bases pour répondre au besoin spécifique des missions.*

Cette approche, bien que pragmatique, soulève des limites méthodologiques, notamment sur le plan de la représentativité des données et de la robustesse des analyses. Jusqu'à présent, l'Igas a fréquemment eu recours à des enquêtes de terrain pour pallier ces lacunes, une solution toutefois contraignante. Les résultats de telles enquêtes nécessitent un contrôle rigoureux pour garantir leur

représentativité et leur redressement peut être chronophage. De plus, cette méthode sollicite les interlocuteurs locaux, déjà confrontés à une charge de travail importante, ce qui peut altérer la qualité et la disponibilité des données recueillies. L'Igas s'efforce désormais d'éviter autant que possible les enquêtes en utilisant des techniques comme le *web scraping*<sup>10</sup> (Lotfi et al., 2021) pour accéder à des données existantes. Cela réduit la sollicitation des acteurs locaux tout en optimisant les analyses avec des données exploitables et dont la fiabilité est acceptable.

<sup>9</sup> Urssaf : union de recouvrement des cotisations de sécurité sociale et d'allocations familiales. L'Agence centrale des organismes de sécurité sociale (Acoss) et le réseau des Urssaf collectent et gèrent les ressources de la majorité des organismes de protection sociale.

<sup>10</sup> Le *web scraping*, ou moissonnage en français, est une technique d'extraction automatisée de données de site web.

À titre d'exemple, dans le cadre de la mission « Lieux de vie et accompagnement des personnes âgées en perte d'autonomie », il convenait absolument de disposer d'une vision globale sur l'offre d'hébergement que proposent les résidences services. Ces dernières sont des structures privées non médicalisées, qui n'obéissent pas au code de l'action sociale et des familles. Elles n'appartiennent donc pas au champ sanitaire et social et ne sont pas recensées dans le répertoire FINESS<sup>11</sup> (Bensoussan et al., 2023). Toutefois, il s'agit d'un acteur important dans l'accueil des personnes âgées, qui tiendra probablement une place de plus en plus grande dans les prochaines années. Pour l'analyse et les projections réalisées, ces résidences devaient donc être prises en compte dans la capacité d'accueil au même titre que les établissements d'hébergement pour personnes âgées dépendantes (Ehpad) ou les résidences autonomie. Les acteurs dans le domaine étant très limités, deux sites web recensaient l'essentiel de l'offre disponible sur le territoire. Le *web scraping* a permis de constituer une base des résidences services ouvertes (ou avec une date d'ouverture prévisionnelle), avec leur lieu d'implantation, les logements disponibles et leur taille, le montant du loyer et les services proposés. Cette base a complété la connaissance sur l'offre d'accueil dans les départements et a été largement utilisée par la mission pour les projections d'accueil des personnes âgées par les différents acteurs du système. L'Igas a bien conscience de la fragilité de ces données, notamment sur les services offerts, dans la mesure où elles s'appuient sur des communications commerciales ; les constats et recommandations en tiennent évidemment compte. Néanmoins, connaître par département cette capacité d'accueil a enrichi substantiellement l'analyse.

### ... ou apprendre à s'en passer

**Le web scraping  
constitue un outil  
intéressant, mais il n'est  
malheureusement  
pas la solution idéale  
au problème de données  
non consolidées.**

Le *web scraping* constitue un outil intéressant, mais il n'est malheureusement pas la solution idéale au problème de données non consolidées. Il suppose de pouvoir répliquer la même requête sur un même site, ou bien de requêter un ensemble de sites de façon similaire. Si les données sont disponibles sur des sites différents, par exemple un site par département dans le cas des résidences services, il est impératif que tous les sites soient construits sur le même modèle. Si la requête se fait sur le même site, ce sont cette fois les outils de sécurité qui sont

potentiellement un obstacle. En effet, une même requête répétée plusieurs fois peut être interprétée comme une cyberattaque et donc être rejetée par le serveur. Ces difficultés conduisent parfois l'Igas à s'adapter pour éclairer tout de même le décideur sur le sujet.

Ainsi, pour la mission « Évaluation de l'encadrement, de l'organisation et de la qualité des vacances adaptées organisées (VAO) », les inspecteurs souhaitaient dans un premier temps dresser un état des lieux des VAO. Une approche sur les sites régionaux ou départementaux aurait pu répondre au besoin, mais tous n'avaient pas l'information. La multiplicité de l'offre ne permettait pas, par ailleurs, une approche par organisme. Des techniques de *web scraping* ont alors été appliquées sur seulement deux organismes

<sup>11</sup> FINESS est le répertoire des établissements sanitaires et sociaux.

bien implantés dans le secteur. L'objectif était d'obtenir l'ensemble des informations sur les séjours proposés aux personnes handicapées (prix, nombre de participants et d'accompagnants), pour les comparer ensuite à leur offre générique. Cette information a permis à la mission d'éclairer concrètement mais partiellement la situation des VAO (Leconte et Itier, 2024).

Le pôle *data* de l'Igas peut être amené à utiliser d'autres techniques d'extraction. Ainsi, le fait que beaucoup d'acteurs du champ social bénéficient d'un financement public, y compris des acteurs privés, se traduit par une place importante du contrôle dans les missions de l'Igas. La coopération des acteurs dans le contrôle ou simplement leur capacité à fournir les éléments utiles pour l'analyse sont alors un élément central des missions. Il peut arriver ainsi que les interlocuteurs transmettent des données numériques sous format PDF, par exemple des liasses fiscales. Or, les outils préformatés de transformation des fichiers ne donnent pas toujours des résultats satisfaisants, ou ils ne respectent pas la confidentialité s'ils doivent transiter par le web pour aller sur un outil d'intelligence artificielle. Dans ce cas, l'Igas a recours à un outil de reconnaissance optique de caractères (ROC ou OCR en anglais pour *optical character recognition*). Un tel outil permet d'extraire les informations importantes et de les stocker dans les fichiers au format ad hoc. Cette technique nécessite un investissement assez coûteux en temps et une structure uniforme des fichiers, mais elle permet de contourner la difficulté. À terme, les outils d'intelligence artificielle installés dans un environnement sécurisé devraient probablement répondre au besoin.

## ► Et maintenant ?

---

Après deux ans d'existence, marqués par l'exploitation de données dont la structure, la qualité, la fraîcheur et la disponibilité restent variables, l'intégration d'une approche *data* dans les missions de l'Igas s'intensifie significativement. Dans ce contexte, il devient impératif pour le pôle d'organiser la capitalisation des travaux déjà réalisés, afin d'en optimiser l'usage et d'en renforcer l'impact.

En premier lieu se pose la question de la maintenance et de l'actualisation des bases de données créées. Le besoin de mise à jour, mais aussi la pertinence et la complétude des données collectées, doivent être évalués. L'évolution du contexte sanitaire et social est un élément majeur qui influence directement les données et dont l'Igas va avoir besoin pour ses analyses. La mise à jour des bases construites par le pôle doit intégrer cette dimension.

### Capitaliser au fur et à mesure des missions

**L'écueil consisterait à répliquer d'une mission à l'autre la même approche *data*, alors qu'il est primordial d'évoluer et de prendre du recul.**

L'Igas réalise régulièrement des missions de contrôle de certains organismes (organismes faisant appel à la générosité publique (OFAG), centres de formation d'apprentis (CFA), etc.) ou d'évaluation d'organismes lors d'un changement de directeur (centres hospitaliers universitaires (CHU), Dreets, etc.). Ces dernières missions sont appelées « T0 ». La récurrence des missions de

contrôle et d'évaluation permet de monter en compétence sur les données, d'améliorer la compréhension qu'en a l'inspection et l'usage qu'elle peut en tirer. Même si le thème de ces missions est identique, chacune doit être perçue comme un nouveau sujet : l'écueil consisterait en effet à répliquer d'une mission à l'autre la même approche *data*, alors qu'il est primordial d'évoluer et de prendre du recul.

Par exemple, dans le cadre du T0 d'un CHU, il est envisageable de systématiser l'extraction de certaines données financières ou d'activité, ainsi que le calcul par grand thème de certains indicateurs identifiés comme pertinents pour comprendre les forces et les voies d'amélioration de l'établissement : évolution des nombres de séances et de patients traités en dialyse ou en chimiothérapie, part de la chirurgie ambulatoire dans l'activité, etc. Cette approche est seulement un premier pas pour orienter des recherches plus poussées sur la situation globale de l'établissement, en prenant par exemple en compte d'autres facteurs comme la concurrence ou les changements de personnel. Il est impératif que les inspecteurs creusent chaque piste afin d'enrichir leur constat et de mettre en œuvre de nouvelles analyses. Ces dernières

pourront à leur tour donner lieu à la définition d'un nouvel indicateur qui sera par la suite calculé systématiquement. C'est pourquoi les *data scientists* doivent échanger avec les inspecteurs en aval des missions, pour construire et enrichir avec eux les travaux qu'ils réalisent.

Enfin, même lorsqu'elles ne sont pas récurrentes, les missions menées par l'Igas répondent aux besoins des décideurs pour orienter leurs politiques et sont en lien direct avec les problématiques de la société. De ce fait, sur une période donnée, plusieurs d'entre elles peuvent

porter sur des problématiques communes. Il est donc important que le pôle sache mettre à profit ses travaux au-delà du champ de la mission elle-même, afin que les missions connexes puissent également en bénéficier. La prise en charge du grand âge qui a donné lieu à plusieurs missions sur les deux dernières années illustre bien cette problématique. Le fait que les analyses quantitatives aient été réalisées par la même équipe constitue un véritable atout. Connaître les bases disponibles et avoir déjà analysé une grande partie des données nécessaires à la mission permet de gagner en efficacité et surtout en qualité de l'analyse. Le *data scientist*, déjà averti sur la problématique, peut avoir une vision plus profonde du sujet et proposer à la mission des pistes d'analyse qu'il n'aurait pas identifiées sans ce travail préalable.

## L'Igas doit oser des opérations innovantes

L'Igas réalise des missions sur un thème précis et ne doit pas nécessairement répondre à des besoins de représentativité nationale. Ainsi, obéissant à moins de contraintes que le système statistique public, le pôle *data* peut réaliser des projets pilotes en collaboration avec les SSM avant que ces derniers n'avancent sur le sujet.

Par exemple, les missions de l'Igas intègrent de plus en plus la notion de parcours. Il peut s'agir de parcours professionnels, comme par exemple dans la mission évoquée plus haut sur les temps partiels contraints, mais aussi de parcours regroupant des thématiques plus variées, nécessitant l'interrogation et la mise en cohérence de plusieurs sources émanant de producteurs différents. L'Igas n'est pas isolée dans cette volonté d'approcher les problématiques des citoyens dans leur globalité. Elle partage ce désir avec ses principaux partenaires de la statistique publique comme la Drees et la Dares, mais ces derniers opèrent dans un cadre plus contraint sur la méthode et doivent s'assurer d'une représentativité à l'échelle nationale.

“

*Le positionnement du pôle data au sein d'une inspection et la connaissance précise qu'il a des sujets portant sur le champ social doivent l'inciter à innover dans ses analyses.*

”

Ainsi, dans le cadre de ses missions, l'Igas s'est associée aux travaux menés par les services de la statistique publique pour étudier la faisabilité d'un croisement de fichiers. L'objectif était de rapprocher les données sur la formation et l'emploi issues de la base Minimas sociaux, droits d'assurance chômage et parcours salariés (Midas) avec celles de l'ASE sur le parcours des enfants passés par leurs services. Ce travail est encore à ses débuts : l'Igas propose d'expérimenter ce rapprochement sur quatre services départementaux d'ASE qu'elle a

étudiés. Un travail entre les producteurs de données, le CASD et l'Igas est en cours. Il pourrait permettre d'étudier localement le devenir professionnel de ces enfants, dont on ne sait encore que très peu de choses. Mais aussi, il pourrait donner un aperçu des difficultés dans la mise en relation des différentes bases. L'avantage d'expérimenter ce rapprochement de sources dans le cadre de la mission est qu'il facilite l'accès direct aux acteurs du terrain, ce qui permet de lever des doutes sur les données locales et d'améliorer la compréhension des parcours. L'analyse de l'Igas ne répondrait pas à l'ensemble des standards et des exigences de la statistique publique, notamment en matière de représentativité. Ce serait une première ébauche, préalable à un travail d'analyse plus complet et représentatif que pourraient faire les SSM.

Il ne s'agit là que d'un cas particulier. Plus largement, le positionnement du pôle *data* au sein d'une inspection et la connaissance précise qu'il a des sujets portant sur le champ social doivent l'inciter à innover dans ses analyses.

## C'est technique, pas magique !

La diversité des missions, la variété des sujets abordés et la richesse des données disponibles font du travail du *data scientist* à l'Igas une expérience à la fois unique et passionnante. Lorsqu'il évolue dans un cadre bien balisé, son travail ressemble à des analyses statistiques classiques. Mais la réalité est souvent plus nuancée : contraintes de temps, qualité des données parfois aléatoire, etc. Le *data scientist* doit alors naviguer avec pragmatisme, s'éloignant parfois des canons de l'orthodoxie statistique pour orienter ses analyses de manière pertinente. Le résultat final, lui, doit rester limpide et accessible à tous, en masquant la complexité et les contorsions intellectuelles qui ont jalonné le processus.

Encore méconnu pour certains, le rôle du *data scientist* pourrait passer, du point de vue des utilisateurs, pour une « moulinette » mystérieuse produisant des résultats sans révéler ses secrets. Pourtant, après deux ans d'existence, le pôle *data* de l'Igas a su démontrer son utilité grâce à une collaboration étroite et constante avec les inspecteurs. Cette coopération a non seulement révélé le potentiel de la *data science* pour les missions de l'Igas, mais aussi mis en lumière ses contraintes et ses limites, en particulier celles liées aux données. Parce qu'en *data science* comme ailleurs, il n'y a pas de magie : juste des données, des algorithmes, du sens critique et du travail.

## ► Bibliographie

---

- BENSOUSSAN, Johanna, BIZINGRE, Joël et COURVALIN, Nathalie, 2023. FINESS, le répertoire des établissements de santé. In : *Courrier des statistiques*. [en ligne]. 11 décembre 2023. Insee. N° N10, pp. 71-92. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7722095?sommaire=7722116>.
- BOURLANGE, Danielle, BRUNET, François, CHIGNARD, Simon et EIDELMAN, Alexis, 2021. Évaluation des besoins de l'État en compétences et expertises en matière de donnée. In : *Rapport de la DINUM et de l'Insee*. [en ligne]. Juin 2021. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.vie-publique.fr/rapport/281669-besoins-de-etat-en-competences-expertises-en-matiere-de-donnees>.
- BRUNET, François, ROTH, Nicole et SCHAPIRA, Irina, 2023. Utilisation des données de la Déclaration sociale nominative (DSN) à des fins de statistiques publiques ou de pilotage. In : *Rapport de l'inspection générale de l'Insee / Rapport de l'Igas*. [en ligne]. Avril 2023. [Consulté le 18 mars 2025]. Disponible à l'adresse : <https://www.igas.gouv.fr/utilisation-des donnees-de-la-declaration-sociale-nominative-dsn-des-fins-de-statistiques-publiques>.
- COTTON, Franck et HAAG, Olivier, 2023. L'intégration des données administratives dans un processus statistique – Industrialiser une phase essentielle. In : *Courrier des statistiques*. [en ligne]. 30 juin 2023. Insee. N° N9, pp. 104-125. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/7635829?sommaire=7635842>.
- COUDIN, Élise et ROBERT, Aude, 2024. Les statistiques sur les causes de décès – Classer et coder... dans la classification internationale des maladies. In : *Courrier des statistiques*. [en ligne]. 16 décembre 2024. Insee. N° N12, pp. 27-50. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/8264550?sommaire=8264562>.
- DIALLO, Cheikh Tidiane, MOREL-JEAN, Constance et SARRON, Clotilde, 2024. L'aide sociale départementale : bénéficiaires, dépenses, financement, personnel – Édition 2024. In : *Les dossiers de la DREES*. [en ligne]. 6 novembre 2024. DREES. N° 124. [Consulté le 4 mars 2025]. Disponible à l'adresse : [https://www.drees.solidarites-sante.gouv.fr/publications-communique-de-presse/les-dossiers-de-la-drees/241106\\_DD\\_aide-sociale-departementale](https://www.drees.solidarites-sante.gouv.fr/publications-communique-de-presse/les-dossiers-de-la-drees/241106_DD_aide-sociale-departementale).
- EMMANUELLI, Julien, FROSSARD, Jean-Baptiste et VINCENT, Bruno, 2024. Lieux de vie et accompagnement des personnes âgées en perte d'autonomie : les défis de la politique domiciliaire, se sentir chez soi où que l'on soit. In : *Rapport de l'Igas*. [en ligne]. 29 mars 2024. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://igas.gouv.fr/Lieux-de-vie-et-accompagnement-des-personnes-agees-en-perte-d-autonomie-les>.
- FOURNIER, Valentine, GROSSE, Alexandre, VILBOEUF, Laurent et VINCENT, Bruno, 2025. Les parcours des usagers de la Sécurité sociale : Comment mieux accompagner les moments importants de la vie ? In : *Rapport de l'Igas*. [en ligne]. 8 avril 2025. [Consulté le 14 avril 2025]. Disponible à l'adresse : <https://www.igas.gouv.fr/les-parcours-des-usagers-de-la-securite-sociale-comment-mieux-accompagner-les-moments-importants-de-la-vie>.

- GADOUCHE, Kamel, 2019. Le Centre d'accès sécurisé aux données (CASD), un service pour la data science et la recherche scientifique. In : *Courrier des statistiques*. [en ligne]. 19 décembre 2019. Insee. N° N3, pp. 76-92. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/4254227?sommaire=4254170>.
- HUMBERT-BOTTIN, Élisabeth, 2018. La déclaration sociale nominative – Nouvelle référence pour les échanges de données sociales des entreprises vers les administrations. In : *Courrier des statistiques*. [en ligne]. 6 décembre 2018. Insee. N° N1, pp. 25-34. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.insee.fr/fr/information/3647025?sommaire=3647035>
- LECONTE, Thierry et ITIER, Christophe, 2024. Vacances organisées pour adultes handicapés : état des lieux et leviers d'amélioration. In : *Rapport de l'Igas*. [en ligne]. 2 juillet 2024. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://igas.gouv.fr/vacances-organisees-pour-adultes-handicapes-etat-des-lieux-et-leviers-damelioration>.
- LOTFI, Chaimaa, SRINIVASAN, Swetha, ERTZ, Myriam et LATROUS, Imen, 2021. Web Scraping Techniques and Applications: A Literature Review. In : *SCRS Conference Proceedings on Intelligent Systems*. [en ligne]. Janvier 2021. Pp. 381-394. [Consulté le 4 mars 2025]. Disponible à l'adresse : [https://www.researchgate.net/publication/367719780\\_Web\\_Scraping\\_Techniques\\_and\\_Applications\\_A\\_Literature\\_Review](https://www.researchgate.net/publication/367719780_Web_Scraping_Techniques_and_Applications_A_Literature_Review).
- MAGNIER, Antoine et VIOSSAT, Louis-Charles, 2024. Temps partiel et temps partiel constraint : des inflexions possibles pour un cadre rénové. In : *Rapport de l'Igas*. [en ligne]. 17 décembre 2024. [Consulté le 4 mars 2025]. Disponible à l'adresse : <https://www.igas.gouv.fr/temps-partiel-et-temps-partiel-constraint-des-inflexions-possibles-pour-un-cadre-renove>.

