



GOVERNEMENT

*Liberté
Égalité
Fraternité*

RAPPORT

Évaluation des besoins de l'État en compétences et expertises en matière de donnée

Danielle BOURLANGE
Inspectrice générale de l'Insee

François BRUNET
Inspecteur général de l'Insee

Simon CHIGNARD
Chargé de mission, Etalab, DINUM

Alexis EIDELMAN
Chef de département, Dares

Sommaire

Synthèse	5
Liste des recommandations	7
Introduction	8
Chapitre 1 : Enjeux	10
1.1 Les données, levier de transformation des politiques publiques.....	10
1.1.1 <i>Les orientations interministérielles, notamment initiées par la DINUM, et leur faible articulation avec la statistique publique</i>	10
1.1.2 <i>Des mesures d'accompagnement</i>	11
1.1.3 <i>Des efforts à poursuivre</i>	12
1.2 Un sujet RH identifié au titre des métiers en tension du numérique.....	13
1.2.1 <i>Un constat partagé mais sans approche spécifique des métiers de la donnée</i>	13
1.2.2 <i>De premières initiatives</i>	14
1.3 Le besoin d'une réponse à la hauteur des enjeux.....	15
Chapitre 2 : Les métiers de la donnée	16
2.1 Qu'est-ce que l'expertise en donnée ?.....	16
2.2 Les différents métiers de la donnée.....	16
2.2.1.1 <i>Data engineer et data architect, des informaticiens experts de la donnée</i>	16
2.2.1.2 <i>Data analyst, data scientist et statisticien, des spécialistes du traitement de données</i>	17
2.2.1.3 <i>Des métiers de gouvernance</i>	18
2.3 Quelles sont les spécificités de ces métiers ?.....	18
2.3.1 <i>Les experts des données, des métiers numériques parmi d'autres ?</i>	18
2.3.1.1 <i>Une indéniable dimension numérique</i>	18
2.3.1.2 <i>... mais avec des spécificités</i>	19
2.3.2 <i>Des prérequis indispensables mais qui peuvent faire défaut</i>	19
2.3.2.1 <i>Les données</i>	19
2.3.2.2 <i>Les infrastructures et l'environnement de travail</i>	20
2.3.2.3 <i>L'acculturation des administrations à la donnée</i>	20
2.4 Des exigences particulières et des risques à prévenir.....	21
2.4.1 <i>La nécessité de comprendre et d'adapter les méthodes au problème posé</i>	22
2.4.2 <i>Les risques à ne pas appréhender les biais des données</i>	22
2.4.3 <i>Un enjeu d'équité incontournable</i>	23
2.4.4 <i>L'isolement du data scientist est un facteur de risque</i>	23
2.5 Une compétence difficile à évaluer.....	24
2.5.1 <i>Au moment du recrutement</i>	24
2.5.2 <i>Et sur les résultats</i>	24

Chapitre 3 : Répondre à la croissance du besoin de compétences.....	26
3.1 Une nette montée en puissance depuis 2017.....	26
3.1.1 <i>Des projets qui se multiplient.....</i>	26
3.1.2 <i>Des recrutements en nette augmentation au cours des dernières années.....</i>	26
3.2 Une enquête qui permet pour la première fois de dresser un tableau des compétences « data » au sein de l'État.....	28
3.2.1 <i>Des besoins fonctionnels couvrant un large spectre.....</i>	29
3.2.2 <i>Des profils très divers pour un effectif global de 2 000 experts.....</i>	29
3.2.3 <i>Hors Insee et SSM, un recours légèrement minoritaire à des fonctionnaires.....</i>	31
3.2.3.1 <i>De forts contrastes dans les profils d'experts.....</i>	31
3.2.3.2 <i>Les data scientists beaucoup moins souvent fonctionnaires.....</i>	31
3.2.3.3 <i>Une différence très significative entre les ministères.....</i>	32
3.2.4 <i>Des équipes qui se structurent de façon relativement diffuse.....</i>	32
3.2.5 <i>Des métiers particulièrement en tension.....</i>	32
3.2.6 <i>Une forte croissance anticipée dans les deux prochaines années.....</i>	33
3.2.7 <i>Une montée en puissance aussi chez les opérateurs.....</i>	34
3.3 Les limites du vivier de fonctionnaires spécialisés sur les métiers de la donnée.....	35
3.3.1 <i>Les corps de l'Insee.....</i>	35
3.3.2 <i>Les autres grands corps.....</i>	36
3.3.3 <i>Les corps techniques.....</i>	37
3.3.4 <i>Une répartition contractuels / prestataires / fonctionnaires subie plutôt qu'organisée.....</i>	38
3.4 Alimenter un vivier de fonctionnaires.....	39
3.4.1 <i>S'appuyer sur les corps de l'Insee pour le métier de data scientist.....</i>	39
3.4.2 <i>Passer par d'autres filières pour les métiers de data analyst et data engineer.....</i>	41
3.5 Activer l'ensemble des leviers dans une approche métier.....	41
Chapitre 4 : Vers une gestion dynamique des compétences.....	43
4.1 Des compétences utilisées mais encore peu pilotées.....	43
4.2 Un enjeu de fidélisation des profils d'experts de la donnée.....	45
4.2.1 <i>Un intérêt à prolonger les contractuels.....</i>	45
4.2.2 <i>Encore trop peu de recours aux contrats à durée indéterminée.....</i>	45
4.2.3 <i>Des difficultés à fidéliser les profils d'experts de la donnée.....</i>	46
4.3 Recommandations.....	47
4.3.1 <i>Traiter la spécificité des métiers de la donnée au sein de la filière numérique.....</i>	47
4.3.2 <i>Développer des processus RH ciblés.....</i>	48
4.3.3 <i>Expérimenter une fonction de référent métier en interministériel.....</i>	49
4.3.4 <i>Développer le travail entre pairs.....</i>	50

Chapitre 5 : Un rôle des services statistiques ministériels repensé.....	52
5.1 Le développement de l'exploitation des données à des fins opérationnelles interroge le rôle des services statistiques ministériels.....	52
5.1.1 <i>Rôle et compétences des services statistiques ministériels.....</i>	52
5.1.1.1 Des missions traditionnelles exercées dans un cadre rigoureux.....	52
5.1.1.2 Pour certains SSM, une expression plus large de leurs missions.....	53
5.1.2 <i>Positionnement des services statistiques ministériels vis-à-vis des besoins d'exploitation des données au sein des ministères.....</i>	54
5.1.2.1 De nouvelles activités émergentes.....	54
5.1.2.2 Une limite éthique et juridique à ces nouvelles activités.....	54
5.1.2.3 Des relations le plus souvent distantes avec les administrateurs ministériels des données.....	55
5.2 Les évolutions à l'œuvre dans les instituts de statistique étrangers.....	56
5.3 Recommandations.....	57
5.3.1 <i>Promouvoir une offre de services.....</i>	58
5.3.2 <i>Clarifier les rôles et développer les collaborations avec les AMD et les services chargés du numérique.....</i>	58
5.3.3 <i>Ouvrir des possibilités de mobilité des experts des SSM vers les pôles ministériels ou interministériel de data science.....</i>	59
Liste des annexes.....	61
Annexe 1 : Lettre de mission.....	62
Annexe 2 : Liste des personnes rencontrées.....	65
Annexe 3 : Bibliographie.....	70
Annexe 4 : Méthodologie du traitement des offres de la place de l'Emploi Public.....	71
Annexe 5 : Questionnaires de la mission.....	72
Annexe 6 : Liste des répondants à l'enquête.....	75
Annexe 7 : Modélisation du nombre de recrutements dans les années à venir.....	76
Annexe 8 : Organisation des services statistiques ministériels.....	78
Annexe 9 : Les compétences « données » des directions du numérique dans les ministères...	79
Annexe 10 : Glossaire.....	81

Synthèse

La donnée est au cœur de la transformation des politiques publiques, comme en témoigne la montée en puissance des initiatives des administrations en matière de big data ou d'intelligence artificielle. La pleine exploitation de ce potentiel implique, pour l'État, de disposer en son sein des compétences et des expertises requises. Cette mission, réalisée à la demande du Directeur interministériel du numérique et du Directeur général de l'Insee, s'est attachée à évaluer qualitativement et quantitativement les **expertises disponibles** au sein de l'État, les **besoins** des administrations à satisfaire à court et à moyen terme et la **capacité à faire face** à l'accroissement de ces besoins.

Ces besoins recouvrent un ensemble de métiers de la donnée caractérisés par leur relative nouveauté et l'absence d'une définition stabilisée. Quatre métiers ont été principalement considérés : *data scientist*, *data analyst*, *data engineer* et *data architect*.

L'enquête menée par la mission auprès des administrations de l'État permet de recenser **environ 2 000 experts de la donnée, dont 700 hors Insee et services statistiques ministériels (SSM)**. Sans être exhaustifs, ces résultats couvrent un champ étendu au sein duquel les directions les plus concernées sont représentées. Ils ne couvrent pas les opérateurs de l'État qui, d'après une collecte partielle, ont aussi des besoins importants. La répartition entre fonctionnaires et contractuels varie fortement, du simple au triple, selon les métiers et encore plus entre ministères.

Les résultats de l'enquête ont permis de projeter les besoins de compétences à un **horizon de deux ans** mais pas au-delà, les administrations ne disposant pas à ce stade d'une vision prospective suffisante. Les besoins supplémentaires ont ainsi été évalués à **400**, portant l'effectif d'experts de la donnée dans le champ de l'enquête à **2 400** en 2023.

La mission a relevé un ensemble de **difficultés susceptibles**, si elles ne font pas l'objet d'une action déterminée, de **remettre en cause la capacité** à faire face aux besoins exprimés. Elle constate tout d'abord l'**insuffisance du vivier de fonctionnaires**, particulièrement pour les emplois les plus experts. Ainsi, alors que près de la moitié des experts de la donnée sont fonctionnaires, ce n'est le cas que de moins d'un quart des *data scientists* (hors Insee et SSM). Parmi les corps de catégories A et A+, seuls ceux de l'Insee (corps des administrateurs et attachés statisticiens) sont spécialisés dans le domaine de la donnée. Toutefois, au regard de leurs effectifs, ils ne sont pas en mesure aujourd'hui d'alimenter de façon significative un vivier de compétences interministériel, en plus de leurs missions relevant du service statistique public.

La **difficulté à fidéliser** les profils en poste, et notamment les contractuels est d'autant plus forte que le marché de l'emploi des experts de la donnée est fortement concurrentiel. L'État employeur ne manque pas d'arguments à faire valoir (notamment l'intérêt des missions, la contribution à l'intérêt général, l'accès à des sources de données uniques), mais il n'est pas toujours en capacité de fidéliser les profils ayant acquis une première expérience, alors même que les administrations pourraient y trouver intérêt. Les contraintes de gestion limitant la revalorisation salariale, le recours encore généralisé au contrat à durée déterminée, le **manque de perspectives de carrières** au sein de l'administration rendent *de facto* le secteur public moins compétitif que le secteur privé auprès des experts de la donnée. Ainsi, l'effort à fournir pour faire croître les effectifs (par recrutement ou reconversion) est d'autant plus important que la fidélisation des profils déjà en poste est incertaine.

La mission relève par ailleurs un **manque de maturité sur les processus RH** en dépit des actions menées depuis 2019 en faveur de la filière numérique. L'évaluation des compétences techniques des candidats est un enjeu d'autant plus important que, sauf exceptions, les équipes de *data science* sont le plus souvent de taille modeste et ne possèdent pas toujours la capacité à évaluer de manière complète et objective les compétences techniques des candidats. La gestion des compétences, des parcours et des carrières est

aujourd'hui rendue compliquée par l'absence de référentiel précis des compétences en matière de données, le manque de pilotage des carrières des experts de la donnée, tant au niveau ministériel qu'interministériel.

Enfin, la mission constate que la **répartition entre fonctionnaires, contractuels et prestataires** est aujourd'hui davantage subie plutôt que réellement organisée. Face aux difficultés rencontrées par les recruteurs et aux limites du vivier de fonctionnaires, la répartition entre les ressources internes et externes, et entre les statuts, est plus souvent le fruit de l'histoire et des pratiques qu'une stratégie délibérée. Or, tous les avis recueillis par la mission convergent sur la nécessité d'un équilibre raisonné entre fonctionnaires, contractuels et prestataires afin de garantir la maîtrise de cette fonction stratégique.

L'articulation avec la statistique publique est encore sujette à progrès, à tous les niveaux. Au niveau ministériel, les administrateurs ministériels des données (AMD) et les services statistiques ministériels (SSM) collaborent peu, voire pas du tout, malgré de réelles perspectives de synergie ; au niveau interministériel, les actions RH menées ciblent quasi-exclusivement **la filière numérique**, alors même qu'une large partie des compétences disponibles au sein de l'État en matière de données se situent **hors de cette filière**.

Au regard des enjeux et de l'ampleur des besoins, il convient dès à présent de mobiliser l'ensemble des leviers disponibles : recrutement de contractuels et de fonctionnaires, fidélisation des profils ayant déjà une expérience au sein de l'administration, meilleure mobilisation des compétences présentes au sein des services statistiques ministériels, et enfin recours à des prestataires, en particulier pour des compétences pointues et qui évoluent très rapidement.

La mission formule **12 recommandations** de nature à répondre à quatre enjeux principaux :

Afin de **répondre à la croissance attendue du besoin de compétences**, elle recommande d'alimenter le vivier de fonctionnaires, notamment en s'appuyant sur les corps de l'Insee pour le métier de *data scientist*, de permettre une meilleure revalorisation des rémunérations, de renforcer la formation continue des agents dans une optique d'acquisition ou de maintien des compétences mais aussi de reconversion professionnelle.

Afin de **créer les conditions d'une gestion dynamique des compétences**, la mission recommande de décliner les actions du plan filière numérique sur les métiers de la donnée, de mettre en place une gestion par les compétences à une maille fine et de fournir aux administrations des outils d'évaluation des compétences techniques. Elle propose d'expérimenter une fonction de référent métier *data scientist* en interministériel.

Afin de **développer le travail et le partage d'expérience entre pairs**, la mission recommande de mettre en réseau les compétences en structurant une fonction d'animation auprès de l'administrateur général des données, des codes sources et des algorithmes ; et de dédier, à titre expérimental, une partie du temps des experts de la donnée à ces travaux d'entraide interministérielle. Cette mise en réseau doit permettre de **gagner en efficacité et de prévenir les risques associés à un exercice isolé de la science des données**.

Enfin, afin de **tirer parti des compétences et des expertises** existantes au sein du **service statistique public**, la mission recommande le développement d'une offre de services des SSM adaptée au contexte de chaque ministère, de formaliser un cadre de collaboration renforcée entre les SSM, les AMD et les services chargés du numérique, et de **faciliter les possibilités de mobilité** pour les experts des SSM en dehors du service statistique public, en privilégiant les pôles de *data science* et l'accès à des postes de responsabilité.

Liste des recommandations

Recommandation 1 : Former à la donnée 50 % des cadres dirigeants d'ici fin 2022, 90 % d'ici 2024.
..... (page 21)

Recommandation 2 : Élargir les recrutements des corps des administrateurs et des attachés statisticiens de l'Insee pour répondre au besoin de *data scientists*.....(page 40)

- Étudier les conditions de cet élargissement : nombre de postes, impact pour les écoles et les cursus de formation...
- Valoriser la mobilité en termes de carrière et favoriser l'accès à des postes de responsabilité en dehors de la statistique publique
- Examiner plus généralement les conséquences de cet élargissement sur l'organisation et le positionnement des corps de l'Insee

Recommandation 3 : Décliner les actions du plan filière numérique sur les métiers de la donnée.(page 47)

Recommandation 4 : Mettre en place une gestion par les compétences, à une maille fine, des experts de la donnée..... (page 48)

- Mettre à disposition des administrations des outils d'évaluation des compétences techniques, en mobilisant le financement du plan de relance

Recommandation 5 : Permettre une meilleure revalorisation des rémunérations des experts de la donnée dans une optique de fidélisation..... (page 48)

- Par exemple en rendant le référentiel des rémunérations applicable au-delà des primo-recrutements

Recommandation 6 : Renforcer la formation continue des agents des administrations aux sujets des données..... (page 49)

- Dans une optique d'acquisition, de maintien des compétences mais aussi de reconversion professionnelle

Recommandation 7 : Expérimenter une fonction de référent métier *data scientist* en interministériel
..... (page 49)

Recommandation 8 : Amplifier la mise en réseau des experts de la donnée en structurant une fonction d'animation auprès de l'administrateur général des données, des algorithmes et des codes sources et en lien étroit avec le directeur général de l'Insee.....(page 50)

Recommandation 9 : Expérimenter en 2022 la possibilité pour les experts de la donnée de consacrer au moins 10 % de leur temps de travail à des projets d'intérêt partagé à une échelle interministérielle
..... (page 51)

Recommandation 10 : Développer une offre de service des services statistiques ministériels adaptée au contexte particulier de chaque ministère.....(page 58)

- Prendre en compte cette offre de service dans la mise en œuvre des feuilles de route ministérielles autour de la donnée, des algorithmes et des codes sources
- Adapter les textes d'organisation pour asseoir un positionnement plus large de ces services en termes d'exploitation de données
- Élaborer une doctrine commune à tous les SSM sur leur champ possible d'intervention au-delà de leurs missions traditionnelles
- Desserrer la contrainte de moyens en lien avec cette offre de service

Recommandation 11 : Formaliser un cadre de collaboration renforcé des services statistiques ministériels avec les AMD et les services chargés du numérique.....(page 59)

Recommandation 12 : Ouvrir davantage les possibilités de mobilité des experts des services statistiques ministériels en dehors du service statistique public, en privilégiant les pôles de *data science* et l'accès à des postes de responsabilité.....(page 60)

Introduction

Les données constituent un formidable levier de transformation de l'action et des politiques publiques. Exploiter l'important patrimoine de données collectées et produites au sein et en dehors de l'administration publique peut en effet permettre de rendre un meilleur service à l'usager, de mieux cibler l'action des administrations ou encore de permettre un pilotage plus précis et agile des politiques publiques.

Face à ces enjeux, les nouvelles techniques de traitement des données massives et d'intelligence artificielle sont porteuses d'importantes promesses. Encore faut-il que les conditions nécessaires soient réunies pour que ces dernières soient tenues. Disposer de compétences expertes, capables d'exploiter ces nouvelles techniques pour élaborer des solutions utiles à l'action publique, est bien entendu une condition première.

C'est cette question centrale des compétences que la mission commanditée par le directeur interministériel du numérique et le directeur général de l'Insee a pour objectif d'éclairer : quels sont les besoins au sein de l'État et comment y répondre ?

Elle s'inscrit dans le contexte d'un renforcement de la politique de la donnée menée par le gouvernement et traduite dans la circulaire du Premier ministre du 27 avril 2021.

La mission a mené une enquête auprès des administrations de l'État et de nombreux entretiens avec les différentes catégories d'acteurs concernés : administrateurs ministériels des données, responsables au sein de services chargés du numérique, de directions métiers et de directions des ressources humaines, gestionnaires des corps de fonctionnaires les plus concernés et responsables de services statistiques ministériels.

Elle s'est attachée à identifier les besoins de compétences par métiers, la propension des administrations à internaliser ou externaliser ces compétences ainsi que les freins rencontrés.

La période récente a vu l'émergence de nouveaux métiers de la donnée aux contours évolutifs et qui ne font pas encore aujourd'hui l'objet d'une classification et de définitions universellement partagées. La mission a centré son analyse sur quatre métiers principaux, définis en adaptant le référentiel de l'Apec au contexte public : les métiers de *data scientist*, *data analyst*, *data engineer* et *data architect*. Le métier de statisticien, le plus représenté au sein des services statistiques ministériels, a été intégré dans l'analyse en l'assimilant en fonction des méthodes et techniques mises en œuvre soit au métier de *data scientist* ou à celui de *data analyst*, en raison des fortes proximités avec ces métiers. La mission n'a en revanche pas considéré le métier de délégué à la protection des données ou *Data protection officer (DPO)* dans la mesure où celui-ci revêt une forte dimension juridique et recouvre des profils de compétences très différents.

Le rapport rappelle en premier lieu les enjeux de la politique de la donnée et les mesures gouvernementales destinées à accompagner le développement de cette politique, qu'il s'agisse de dispositifs ciblés sur le domaine des données et de l'intelligence artificielle ou de dispositifs concernant plus généralement la filière et la transformation numériques des administrations de l'État (chapitre 1).

Le rapport dresse ensuite un panorama des différents métiers de la donnée retenus par la mission, en soulignant leurs particularités ainsi que les exigences et risques spécifiques qui leur sont attachés, la puissance des outils et le volume de données disponibles ne suffisant pas à garantir la pertinence des résultats obtenus. Il présente à ce titre les prérequis indispensables pour une mise en œuvre efficiente de ces métiers, tant en termes de qualité des données, d'infrastructures techniques que de nécessaire acculturation de l'encadrement ainsi que les risques à considérer les *data scientists*, sous le seul angle technique (chapitre 2).

Le chapitre 3 présente les résultats de l'enquête menée par la mission qui permet pour la première fois de dresser un état des lieux des compétences *data* au sein de l'État par métier et d'en préfigurer l'évolution dans les deux prochaines années. Ces résultats sont mis en perspective avec la capacité des corps de l'État à alimenter un vivier de fonctionnaires spécialisés dans les métiers de la donnée et conduisent la mission à formuler des recommandations pour améliorer la réponse au besoin de compétences.

Le chapitre 4 met en lumière la nécessité de développer une gestion dynamique des compétences avec une attention particulière portée à la formation continue, à l'évaluation et au suivi des compétences, aux possibilités de mobilité et de parcours et à une réponse ciblée au besoin de fidélisation des contractuels. La mission considère que les mesures du plan d'action en faveur de la filière numérique portées par la DINUM doivent à ce titre faire l'objet d'une déclinaison spécifique pour les métiers de la donnée et recommande des mesures complémentaires ciblées.

Le chapitre 5 examine la place qu'occupent les services statistiques ministériels au sein des écosystèmes ministériels de la donnée et la contribution qu'ils pourraient apporter aux besoins de compétences des directions de leurs ministères en exploitant davantage les synergies entre les expertises développées dans le cadre des travaux de statistique publique et celles qui sont nécessaires à une exploitation des données à des fins opérationnelles. Des recommandations sont formulées dans ce sens ainsi qu'en faveur d'une collaboration plus étroite des services statistiques ministériels avec les administrateurs ministériels des données et les services chargés du numérique.

Chapitre 1 : Enjeux

1.1 Les données, levier de transformation des politiques publiques

L'État n'a pas attendu l'émergence des données massives pour se préoccuper des données. Les données aujourd'hui mobilisables sont multiples et diverses : données d'enquête et données administratives, données géographiques, données de transaction voire même traces numériques (à l'instar des données anonymisées de téléphonie mobile).

La puissance publique est depuis longtemps productrice de données, pour ses besoins propres et pour ceux de la société¹. La production de données de référence s'appuie sur les grands producteurs que constituent l'Insee, l'Institut national géographique et forestier ou la Direction générale des finances publiques. Ces données de référence, reconnues par la loi, servent de base pour nombre de procédures et de services en ligne du secteur public mais aussi du secteur privé. Le service statistique public, composé de l'Insee et des 16 services statistiques ministériels (SSM), a pour mission de collecter les données nécessaires à l'élaboration de statistiques, de produire et de diffuser auprès d'un large public les données officielles sur lesquelles s'élabore le débat public, de fournir des analyses chiffrées (cf. chapitre 5).

La circulation des données est l'un des piliers de la politique publique de la donnée, avec la mise en œuvre du principe « dites-le nous une fois » et l'ouverture des données publiques et des API. Une meilleure circulation de la donnée, au sein et en dehors des administrations, est un pré-requis à la pleine exploitation du potentiel des données. Cette exploitation a connu une accélération depuis le milieu des années 2010, notamment en raison de l'émergence de nouvelles techniques d'analyses (apprentissage machine) mais aussi de l'accroissement du volume et de la diversité des données traitées (données administratives, données de transaction, traces numériques, etc.).

Les données contribuent à la transformation de l'action et des politiques publiques. Elles permettent notamment de rendre un meilleur service à l'utilisateur (à l'image des déclarations d'impôts pré-remplies), de mieux cibler les actions des administrations (par exemple dans le domaine du contrôle fiscal) ou encore de permettre un pilotage plus précis et agile des politiques publiques (comme l'illustre le tableau de bord de suivi des aides économiques liées à la crise de la Covid-19).

1.1.1 Les orientations interministérielles, notamment initiées par la DINUM, et leur faible articulation avec la statistique publique

La gouvernance de la donnée de l'État relève aujourd'hui essentiellement de la Direction interministérielle du numérique (DINUM). Cette gouvernance est aujourd'hui faiblement articulée avec celle de la statistique publique, tant au niveau interministériel qu'au sein des ministères.

Le décret du 16 septembre 2014 institue, auprès du Premier ministre, un Administrateur général des données (AGD). L'AGD coordonne l'action des administrations en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données. Il organise, dans le respect de la protection des données personnelles et des secrets protégés par la loi, la meilleure exploitation de ces données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation.

Le positionnement de la fonction d'AGD évolue au cours des années 2015-2019 : initialement placé auprès du Secrétariat général pour la modernisation de l'action publique (disparu en 2017), il est ensuite rattaché à la Direction interministérielle du numérique. Le décret du 25 octobre 2019 relatif au système d'information et de communication de l'État et à la direction interministérielle du numérique précise le rôle de cette dernière en matière de données. La DINUM « *coordonne et promeut l'action de l'État et des organismes placés sous*

¹ Administrateur général des données : rapport au Premier ministre sur la gouvernance de la donnée, les données au service de la transformation de l'action publique, La Documentation française, 15 décembre 2015

sa tutelle en matière d'inventaire, de gouvernance, de production, de circulation, d'exploitation et d'ouverture des données ». Le directeur interministériel du numérique exerce, à ce titre, la fonction d'administrateur général des données².

L'organisation en matière de système d'information et de communication de l'État est plus généralement définie par le décret n°2019-1088 du 25 octobre 2019. Les ministères sont tenus de mettre en place des directions du numérique qui ont un rôle en matière de données : « *Chaque ministère met en place une structure qui organise et pilote les actions du ministère en matière de numérique, et à ce titre : l'exploitation du potentiel offert par les données* ».

L'organisation qui résulte du décret du 25 octobre 2019 se caractérise à la fois par l'intégration de la donnée comme l'une des missions parmi d'autres de la DINUM (et des directions ministérielles du numérique) et par l'absence d'articulation avec les services statistiques ministériels. De ce point de vue, la gouvernance de la donnée publique apparaît comme un sujet disjoint de celui de la statistique publique, quand bien même l'Insee et les services statistiques ministériels (SSM) constituent des pôles reconnus en matière de production et d'exploitation de données. De manière générale, les textes d'organisation des directions ministérielles du numérique ne mentionnent pas l'articulation avec les services statistiques ministériels. Il convient enfin de noter que le mot même de « données » est, sauf exceptions, absent des textes statutaires organisant les missions des SSM (c'est notamment le cas pour la Drees, la Dares, le SSP agriculture, la DEPP, le SSM culture, le SDES³).

La fonction d'administrateur ministériel des données (AMD) n'avait jusqu'à une date récente fait l'objet d'aucun texte réglementaire permettant de l'encadrer et d'en définir les missions. Le département Etalab se charge, depuis 2016, de l'animation d'un réseau interministériel des administrateurs ministériels des données. Il convient de noter qu'à une exception près, les SSM ne jouent pas le rôle d'AMD (cf. chapitre 5).

1.1.2 Des mesures d'accompagnement

La DINUM porte, seule ou en collaboration avec la Direction interministérielle de la transformation publique (DITP), un ensemble de dispositifs visant à une meilleure exploitation des données au sein des administrations : programme des Entrepreneurs d'intérêt général (EIG), action « pilotage par la donnée » (DA6) du programme TECH. GOUV, appels à manifestation d'intérêt sur l'intelligence artificielle (AMI IA), fonds de transformation de l'action publique (FTAP) et, plus récemment encore appels à projet dans le cadre du plan de relance (France Relance).

Le programme EIG, lancé en 2016, part du constat que « *les administrations qui veulent mener des projets de transformation numérique ne disposent pas des compétences pointues nécessaires en interne pour les réaliser* »⁴. En réponse, le programme EIG propose à des profils numériques d'intégrer des administrations afin de mener sur un temps court (dix mois) des projets techniques présentés sous la forme de défis. Pour ce faire, le programme constitue chaque année une promotion d'entrepreneurs d'intérêt général composée de profils divers (*data scientists, designers, développeurs...*). Au cours des quatre premières éditions du programme EIG, 70 projets (ou défis, selon la terminologie retenue par le programme) issus de 40 administrations ont été accompagnés par plus de 130 entrepreneurs d'intérêt général.

Les projets, proposés par les administrations et sélectionnés par le programme, sont fortement marqués par la thématique *data*. Le programme EIG vise à la fois la réalisation d'outils opérationnels, mais aussi plus largement à introduire au sein des administrations des profils et compétences nouvelles (dont les *data scientists* et les *designers*). Il convient de noter qu'une attention particulière est portée aux projets des administrations qui prévoient l'intégration, tant des compétences que des solutions développées, dans l'administration à l'issue de la phase de développement. Les conditions de généralisation et de pérennisation des projets font aujourd'hui partie des critères d'évaluation.

² Décret n°2019-1088 du 25 octobre 2019 relatif au système d'information et de communication de l'État et à la direction interministérielle du numérique, article 6.

³ Voir annexe 7

⁴ Présentation du programme EIG sur le site entrepreneur-interet-general.etalab.gouv.fr

L'appel à manifestation d'intérêt IA (AMI IA) a été lancé 2018 conjointement par la Direction interministérielle de la transformation publique (DITP) et la Direction interministérielle du numérique (DINUM) dans le cadre du programme d'investissements d'avenir. L'enjeu est d'expérimenter concrètement l'usage des technologies d'intelligence artificielle pour répondre à un besoin métier d'une administration. Ces dernières sont invitées à candidater et les projets retenus bénéficient d'un financement et d'un accompagnement dédié par la DINUM et la DITP ainsi que le support de l'Inria. Le programme est monté en puissance depuis son lancement : la première édition a accompagné six projets et la seconde quinze projets. L'AMI IA partage avec le programme EIG une temporalité courte (dix mois) et un processus de sélection des projets accompagnés. Cependant, contrairement au programme EIG, l'appel à manifestation d'intérêt IA ne vise pas explicitement à introduire de nouvelles compétences au sein des administrations lauréates.

Dans le cadre du plan de relance, plusieurs programmes de financement ont été fléchés sur des thématiques en lien avec les données. Le Ministère de la Transformation et de la Fonction publique porte ainsi le volet « mise à niveau numérique de l'État et des territoires » dont une sous-partie est consacrée au développement de l'utilisation de la donnée dans l'administration (ITN5) et à la professionnalisation des filières numériques publiques (ITN4). Le premier appel à projets, clôturé fin 2020, a permis de sélectionner une quinzaine de projets portés par des administrations. Ces projets concernent tout le cycle de vie de la donnée, de la production à l'exploitation. Par ailleurs, d'autres ministères sont susceptibles de porter des actions interministérielles sur le sujet, à l'instar du marché interministériel de prestations de service porté par le Ministère de l'Intérieur, dont l'un des lots vise explicitement la *data science* et l'intelligence artificielle.

1.1.3 Des efforts à poursuivre

Le rapport « Pour une politique publique de la donnée », remis au Premier ministre par le Député Eric Bothorel le 23 décembre 2020, encourage l'État à poursuivre et accélérer ses efforts en matière d'exploitation des données, tant issues du secteur public que du secteur privé (données d'intérêt général).

Parmi les freins à une meilleure exploitation des données, la question des talents est centrale : « *la gestion des compétences et les outils RH pour attirer et maintenir les profils spécialisés progressent, mais ne sont pas encore suffisants pour garantir une filière technique de haut niveau dans le domaine de la donnée et du code* » indiquent les auteurs. Le rapport Bothorel recommande en outre de « *diversifier les parcours des administrateurs et des attachés de l'Insee dans l'ensemble des administrations, au-delà des services statistiques ministériels* ». Enfin, le rapport appelle à développer une culture de la donnée au sein des administrations, en particulier par le biais d'une politique de formation ambitieuse et qui concernerait tous les échelons administratifs (le rapport mentionne notamment une « obligation de formation des cadres dirigeants aux enjeux du numérique »).

Le 5e Comité interministériel de la Transformation publique (CITP) du 5 février 2021 a réaffirmé l'objectif, dans la lignée de la remise du rapport « Pour une politique publique de la donnée » au Premier ministre : « *une politique publique de la donnée ambitieuse sera mise en œuvre par tous les ministères pour mieux exploiter, partager et valoriser les données publiques (...)* ». Cet engagement est décliné sur plusieurs volets, dont celui des ressources humaines et des compétences.

La circulaire du Premier ministre du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources indique que « *la politique de la donnée doit constituer une priorité stratégique de l'État (...)* ». Le calendrier de mise en œuvre précise que les ministères devront avoir, au 15 mai 2021, nommé un administrateur ministériel des données et devront fournir, d'ici le 15 juillet, une feuille de route présentant la stratégie ministérielle dans le domaine de la donnée, des algorithmes et des codes sources. Il est enfin précisé que les dites feuilles de route doivent aussi veiller au développement des « *compétences relatives aux données, algorithmes et codes sources (...), notamment pour les cadres dirigeants, supérieurs et intermédiaires* ». Enfin, la circulaire du Premier ministre évoque aussi la stratégie interministérielle pour la filière numérique incluant les compétences liées aux données et aux algorithmes.

1.2 Un sujet RH identifié au titre des métiers en tension du numérique

1.2.1 Un constat partagé mais sans approche spécifique des métiers de la donnée

Le rapport sur la stratégie de ressources humaines de la filière numérique et systèmes d'information et communication de l'État de janvier 2019 dresse un bilan des enjeux RH⁵. « *Longtemps considéré comme une simple fonction support, la fonction SIC (système d'information et de communication) est devenue une mission stratégique, aussi bien pour les entreprises que pour l'État, a fortiori dans une période où le Gouvernement veut faciliter la transformation de l'action publique en s'appuyant sur le numérique* » écrivent les auteurs. Ce rapport fait état d'un consensus autour des métiers en tension. La définition d'un métier en tension, retenue par la DGAFP et la DINSIC en 2017, est que le « *taux d'emploi vacant est supérieur à la moyenne de la filière, et le fait que les effectifs soient vieillissants* ». Treize métiers sont ainsi identifiés dans la filière numérique, soit en raison de leur rareté, soit en raison de leur caractère stratégique. Le métier de *data scientist* est l'un de ces métiers en tension et il est qualifié de critique au sens où il combine les deux dimensions (rareté et caractère stratégique). D'après l'enquête menée en 2017, les besoins évalués pour l'année 2020 dépassent de 1,6 fois le nombre d'agents actuellement en poste.

Les auteurs du rapport de 2019 apportent toutefois une nuance : « *il faut tenir compte de l'effet de mode toujours très présent dans le monde de l'informatique, qui fait que les métiers nouveaux sont par définition peu représentés dans les effectifs actuels et forts demandés, ce qui conduit à des taux de vacances élevés car appliqués à des effectifs faibles* ». Le métier de *data scientist* est relativement nouveau par rapport aux autres métiers en tension (architecte technique, urbaniste des systèmes d'information et, dans une moindre mesure, auditeur en sécurité des systèmes d'information). Le rapport formule un ensemble de recommandations sur le recrutement de contractuels en CDI, le renforcement des effectifs du corps des ISIC ou encore la formation des agents. Il appelle aussi à renforcer l'attractivité de l'État employeur, soulignant aussi le caractère concurrentiel du marché de l'emploi : « *Si les rémunérations en début de carrière paraissent compétitives par rapport au marché, l'écart se creuse ensuite en défaveur de l'État qui est alors pénalisé pour ses recrutements en milieu de carrière. Comme il ne sera pas possible de s'aligner sur le haut du marché, il faut mettre en place d'autres mécanismes qui permettent de motiver et de conserver les agents en poste* ».

Le sujet RH des métiers de la donnée est donc bien identifié, mais d'une part il se limite uniquement au profil de *data scientist*, et d'autre part il est principalement traité sous l'angle des métiers en tension. Aucune mesure ne concerne donc spécifiquement ces métiers de la donnée.

Le plan d'actions détaillé par la circulaire DGAFP-DINUM du 2 mai 2019 répond à trois objectifs :

- attirer et recruter les bons profils dans le vivier spécifique des agents de la filière NSIC ;
- développer la mobilité et les parcours professionnels des agents de la filière ;
- faciliter le recours au contrat pour les métiers NSIC.

La circulaire mentionne ainsi 31 actions ainsi qu'un calendrier de mise en œuvre de 2019 à 2021. Chaque action se voit attribuer un niveau de priorité, de 1 à 4. On notera en particulier la mise en place d'un ensemble d'outils : kit de communication autour de la marque de l'État-employeur (priorité niveau 1, 2019), outil de gestion prévisionnelle des emplois, des effectifs et des compétences (priorité niveau 4, 2021). Enfin, le plan d'actions prévoit la mise en place d'un référentiel de rémunération pour les métiers NSIC (priorité niveau 1, 2019-2020) ainsi que l'accompagnement de l'ensemble des parties prenantes (DSI, DRH, CBCM) dans l'utilisation de ce référentiel. La première version de ce référentiel concerne prioritairement les quinze métiers en tension.

⁵ Rapport sur la stratégie de ressources humaines de la filière numérique et Système d'Information et de communication de l'État, inspection générale de l'administration, conseil général de l'économie et contrôle général économique et financier, janvier 2019

Il est enfin notable de constater que ni le rapport du CGE de 2019, ni la circulaire DGAFP-DINUM de la même année ne mentionnent la possibilité que certains besoins de la filière NSIC, et en particulier le profil de *data scientist*, puissent être remplis par des corps qui n'appartiennent pas à la filière numérique. Par exemple, il n'est pas fait mention des corps de l'Insee (administrateurs et attachés) et de leur capacité à fournir des profils d'experts de la donnée.

1.2.2 De premières initiatives

En application de la circulaire DGAFP-DINUM de 2019, plusieurs initiatives ont vu le jour notamment autour de la marque employeur de la filière numérique, de la mise à jour du référentiel des métiers et compétences SIC, de la mise en œuvre d'un référentiel de rémunération pour les métiers en tension ou encore dans le domaine de la formation aux compétences numériques. La mission Talents, au sein du programme TECH. GOUV de la DINUM est chargée de l'animation des travaux interministériels, en lien avec la DGAFP et les ministères.

Le travail sur la marque employeur se décline à plusieurs niveaux⁶. Conformément au plan d'actions de la circulaire de 2019, un kit de communication a été mis en place pour la filière numérique à l'été 2019. L'objectif de cet outil est de « *faire connaître la filière NSIC dans la fonction publique, pour mieux attirer et recruter les profils recherchés. Il constitue la première brique des travaux engagés en vue de développer la marque État Employeur dans cette filière* ». Ce kit, élaboré par la DINUM, a vocation à être utilisé par les représentants des ministères lors de leur participation à des forums, salons ou colloques relatifs à l'emploi public. Il se décline notamment autour de profils d'agents (« personæ ») établis selon leur métier, leur formation, leur catégorie d'âge. À chaque profil est associé des leviers d'attractivité (service public, intérêt des projets, conditions et qualité de vie au travail...). Il convient de noter que la spécificité de cette marque employeur, qui s'attache à traiter d'une filière particulière, celle du numérique. Cela vient en complément des initiatives existantes, soit au niveau ministériel (Ministère des Armées ou de la Justice), soit en interministériel. S'agissant de la communication externe, il est prévu que la marque employeur de la filière numérique s'inscrive comme une déclinaison de la marque employeur de l'État sur laquelle travaille la DGAFP et l'APIE pour l'ensemble de la fonction publique. Une déclinaison dans les campagnes de communication est prévue courant 2021. Les résultats des travaux sur le sujet (diagnostics, stratégie, éléments de largage) sont partagés avec les ministères à chaque étape afin qu'ils puissent s'en inspirer dans leurs campagnes de communication ministérielle ainsi qu'à l'occasion des événements interministériels.

Une logique similaire de déclinaison, pour la filière numérique, d'outils existants au niveau de l'État a été mise en place pour le recensement et la définition des métiers. Le Répertoire interministériel des métiers de l'État (Rime) définit près de 300 emplois-références, dont une vingtaine dans le domaine du numérique et du système d'information et de communication (SIC). À compter de 2017, la DINUM a élaboré un référentiel des métiers et compétences SIC. À ce jour, ce référentiel comporte près d'une soixantaine de métiers, dont une dizaine concerne directement la donnée, la mise à jour la plus notable datant de 2019.

Le référentiel de rémunérations des quinze métiers en tension de la filière numérique et système d'information et communication (NSIC) est présenté comme l'un des principaux outils pour permettre le recrutement de contractuels aux compétences rares ou pointues. Le document publié en juillet 2019 est signé par la DGAFP, la DINUM et la Direction du Budget et a pour ambition de « *simplifier le recrutement sur ces métiers en tension en réduisant le champ des visas des CBCM et de mettre en cohérence les pratiques de gestion, de limiter la concurrence entre les ministères* ». Il convient de noter que ce référentiel cible le recrutement de nouveaux contractuels au sein de l'État et qu'il ne concerne donc pas les renouvellements de contrats. Pour chaque métier en tension, le référentiel fait apparaître des plafonds de rémunération brute annuelle permettant de dispenser le recruteur d'un visa du CBCM. Ce plafond varie en fonction du métier et du nombre d'années d'expérience. Les montants sont issus d'une étude des rémunérations dans le secteur privé.

⁶ « Transformation de la fonction recrutement, éléments préparatoires au plan interministériel de transformation de la fonction recrutement, édition 2020 », DGAFP

À titre d'illustration, pour le profil de *data scientist* (le seul métier lié aux données dans les quinze métiers en tension), le plafond de rémunération est de 50 000 € (pour les profils ayant moins de cinq ans d'expérience), 59 000 € (entre cinq et dix ans d'expérience) et 80 000 € (plus de dix ans d'expérience).

La DINUM indique qu'une mise à jour de ce référentiel de rémunération, attendue pour le second semestre 2021, doit intégrer l'ensemble des métiers du référentiel des métiers, au-delà des quinze métiers en tension.

Dans le domaine de la formation aux compétences numériques, un marché interministériel, porté par le Ministère des Armées avec la DGAFP et la DAE (direction des achats de l'État) a été notifié en juillet 2020. Ce marché compte huit ministères adhérents et 54 établissements publics. 700 formations sont destinées à près de 10 000 stagiaires dans le domaine du numérique et des systèmes d'information et de communication. L'un des lots du marché (lot 5 – développement, gestion et analyse de données) comprend une dimension Big Data et Intelligence artificielle.

Enfin, une mise à jour de la circulaire DGAFP-DINUM de 2019 a été annoncée en Comité interministériel du numérique (CINUM) en mai 2021. Cette mise à jour comporte une vingtaine d'actions, mais aucune ne cible de manière spécifique les métiers de la donnée.

1.3 Le besoin d'une réponse à la hauteur des enjeux

Au-delà des initiatives déjà lancées, qui ne visent pas spécifiquement le domaine des données, la question posée à la mission est bien celle des leviers à actionner et des mesures à prendre afin de favoriser la meilleure réponse aux besoins de compétences, en tenant compte des spécificités des métiers de la donnée et des exigences qui leur sont attachées (cf. chapitre 2).

Faute de ressources et de compétences à niveau pour exploiter les données, la puissance publique risque en effet de se priver d'un levier majeur de transformation et d'efficacité. Il n'y a pas de politique publique de la donnée sans les compétences nécessaires pour produire, traiter et exploiter les données, tant du secteur public que privé (données d'intérêt général). Faute de compétences internalisées et à jour, l'État prend aussi le risque de se voir déposséder d'une capacité à faire et à réguler, en particulier face à des grands acteurs du numérique. Ces derniers se caractérisent, outre leur capacité à collecter massivement des données, par leur savoir-faire en matière de *data science*. Si la puissance publique se désarme, elle ne peut pas prétendre tenir son rôle, par exemple dans le domaine des données de santé ou d'environnement. Avec l'idée que l'intelligence artificielle et les données en général peuvent permettre des gains d'efficacité, ne pas se donner les moyens de les réaliser, c'est risquer de ne pas dégager des marges de manœuvre.

Pour certains domaines de l'action publique, il ne s'agit pas seulement d'efficacité, c'est l'accomplissement même de la mission qui est en jeu. On peut penser au domaine militaire où il s'agit de maîtriser l'intelligence artificielle aussi bien ou mieux que les autres pays dans une problématique d'équilibre des forces, mais aussi aux domaines de la régulation. En effet pour la régulation concurrentielle des marchés, il est nécessaire que la puissance publique dispose de compétences au niveau de celles des entreprises contrôlées.

Le sujet du contrôle des algorithmes est récent mais va aller croissant : comment mesurer la position dominante d'un acteur si on ne peut pas connaître son algorithme, comme évaluer la non-discrimination sans maîtrise des données ? Comment déterminer des règles pour faire en sorte que les prix fixés par des algorithmes ne reproduisent pas, parfois malgré eux, des systèmes illégaux de cartels de prix ? Sans connaissance avancée du fonctionnement de ces algorithmes, il n'est pas possible de comprendre les mécanismes à l'œuvre, de les contrôler, et de proposer des mécanismes de régulation.

Chapitre 2 : Les métiers de la donnée

2.1 Qu'est-ce que l'expertise en donnée ?

La mission définit comme étant « experts de la donnée » les personnes dont les données sont l'objet essentiel du travail et non seulement un moyen de le réaliser. Cela inclut les métiers qui assurent le traitement informatique des données (et non seulement la collecte par une interface) et ceux qui exploitent les données pour produire des analyses, des recommandations, des prédictions.

Tous ces experts ont en commun de savoir que les données sont le résultat d'un processus de captation ou de collecte et que celui-ci n'est ni neutre ni anodin. Chaque capteur, chaque mode de collecte, chaque source de données a ses spécificités et ses implications, tant au niveau des infrastructures et des traitements informatiques que du point de vue de l'analyse que l'on peut en tirer. Les experts de la donnée savent aussi adapter les données à l'usage qui en est fait. L'infrastructure, le traitement, l'analyse ou la présentation vont se faire différemment selon le besoin. L'architecture des données et les traitements informatiques à mettre en place sont par exemple différents selon la réactivité désirée, l'exigence de sécurité et l'usage qui est fait des données. Les opérations de mise en qualité des données, les corrections, imputations et estimations ne sont pas les mêmes selon l'objectif du traitement. Enfin, la présentation des données et des analyses dépend aussi des destinataires.

La mission définit deux métiers à dominante informatique : le *data engineer*⁷ et le *data architect* et deux métiers à dominantes mathématiques / statistiques : *data analyst* et *data scientist*. Le statisticien est également un expert de la donnée et la proximité de ce métier avec les précédents sera précisée dans ce chapitre. La mission a néanmoins choisi de ne pas interroger les directions sur ce métier en recommandant d'intégrer la plupart des statisticiens aux catégories *data analyst* et *data scientist*, dans une approche qui met l'accent sur les compétences techniques mises en œuvre et de façon à distinguer au mieux les *data scientists*.

La mission a également fait le choix de ne pas traiter les experts juridiques de la donnée et en particulier les délégués à la protection des données dont l'expertise technique est bien différente de celle des premiers métiers et qui mériterait d'être l'objet d'une mission à part. Enfin, compte tenu de l'émergence des métiers et de la faiblesse supposée de l'effectif concerné, la mission n'a pas non plus intégré dans son recensement les métiers d'encadrement, supérieur ou non, chargés de la gouvernance des données ou *chief data officer* ni le métier de *data steward*.

2.2 Les différents métiers de la donnée

Du fait de leur nouveauté et peut-être de leur relative technicité, les métiers de la donnée ne sont pas toujours bien connus et, malgré les divers travaux de définition (Apec, Cigref, etc), le contour des métiers est en évolution constante. Le Répertoire Interministériel des Métiers de l'État (Rime), dans la dernière version au moment de la rédaction du rapport, comprend ainsi seulement deux catégories de métiers de la donnée : le *data scientist* et le chargé de gouvernance de la donnée, là où la mission en définit quatre ; à l'inverse le guide « *Data scientist !* » édité par Télécom Paris en considère une vingtaine.

2.2.1.1 *Data engineer* et *data architect*, des informaticiens experts de la donnée

Le *data architect* préconise les solutions techniques à mettre en œuvre dans les projets nécessitant la collecte, le stockage ou l'utilisation de données en masse. Il met en place des solutions génériques pour faciliter le travail des équipes de développeurs.

⁷ La mission a choisi d'utiliser les termes en langue anglaise. C'est en effet de cette manière que ces métiers sont les plus souvent désignés.

À la fois technicien qui travaille avec les *data engineers* mais aussi avec les développeurs applicatifs et le *software engineer* pour définir la meilleure plateforme et programmer les mises à jour, il doit également savoir communiquer avec de nombreuses parties prenantes, dont les *data scientists*. Il aide à définir des objectifs stratégiques en matière de gestion de données. Devant se tenir à jour des innovations dans le secteur, le *data architect* doit aujourd'hui pouvoir préconiser et utiliser les solutions *cloud* à bon escient. Même si, en nombre, le besoin est moins important que pour les autres métiers de la donnée, c'est *a priori* le métier le plus en tension au moment de la rédaction de ce rapport.

Le *data engineer* organise, en tant que développeur informatique, la collecte et la mise à disposition des données au sein du ministère ou de la direction. Il industrialise et met en production des traitements sur les données (par exemple : mise à disposition de tableaux de bord, intégration de modèles de *machine learning*) en lien avec les équipes métiers et les équipes de *data science*. Il assure le lien entre la plateforme déterminée par le *data architect* et le besoin de données exprimé par l'utilisateur de données qu'il soit *data analyst*, *data scientist* ou statisticien. Il intervient également en aval du travail du *data scientist* dans les étapes d'implémentations opérationnelles des algorithmes élaborés par celui-ci.

2.2.1.2 *Data analyst, data scientist et statisticien, des spécialistes du traitement de données*

Le *data analyst*, ou analyste des données, examine des données bien définies dans le but de produire des informations pertinentes dans un contexte métier. Sa boîte à outils statistiques et informatiques lui permet d'organiser, synthétiser et traduire les informations utiles aux organisations pour orienter les prises de position des acteurs décisionnels. Dans certains contextes, ces informations sont appelées indicateurs clefs de performance pertinents, parfois outils de pilotage ou encore simplement statistique. Il peut réaliser des traitements sur des logiciels intégrant une partie de programmation et une capacité à produire et à actualiser des représentations graphiques (Excel VBA, Tableau, Qlick, etc.). Œuvrant plus à la restitution de données qu'au développement de modèles prédictifs, il peut évoluer vers le métier de *data scientist* ou le métier de statisticien dont il est proche.

Le *data scientist*, ou scientifique des données ou ingénieur en intelligence artificielle, effectue des tâches complexes dans le traitement des données. Il est capable de traiter des données variées et de mettre en place des algorithmes optimisés de classification, de prédiction sur des données numériques, textuelles ou d'images. Les données qu'il traite peuvent être structurées, non-structurées ou intermédiaires. Au sein de ce métier, des spécialités peuvent apparaître selon le type de donnée ou la technologie utilisée. Le *data scientist* utilise des outils de programmation et doit savoir optimiser ses calculs pour les faire tourner rapidement en exploitant au mieux les capacités informatiques (serveur local, *cloud*, CPU, GPU, etc.) Positionné sur un métier en évolution rapide, il doit être capable de suivre l'avancée des technologies et de se former régulièrement.

La distinction entre le *data scientist* et le **statisticien** traditionnel est réelle mais sans être incommensurable. La proximité entre ces deux métiers est d'ailleurs telle que certains services statistiques enquêtés ont considéré que l'ensemble des statisticiens étaient des *data scientists*. Cela correspond à la définition informelle qui décrit le *data scientist* comme « un statisticien qui s'y connaît en informatique » (ou un informaticien qui s'y connaît en statistique). Même si la finalité, l'environnement des données et les environnements techniques diffèrent parfois, les concepts mathématiques mis en œuvre, la disposition d'esprit nécessaire pour exercer ces métiers sont très proches. Au niveau technique, malgré la nouveauté de l'efficacité de certaines méthodes (réseaux de neurone, forêts aléatoire), une grande partie des méthodes mises œuvre par le *data scientist* sont employées de longue date par les statisticiens (régression logistique, classification ascendant hiérarchique, etc.). Les deux profils ont en commun d'interroger la méthode employée pour savoir si elle est valable. Du reste, même si des formations spécifiques se développent par exemple dans les écoles d'ingénieurs, les écoles de formations aux statistiques sont également des écoles de formation à la *data science*.

Enfin ces métiers requièrent, au-delà de solides compétences techniques, une connaissance de l'environnement relatif aux données. Il ne s'agit pas simplement d'appliquer des méthodes, il faut une « connaissance métier ». Celle-ci est indispensable. Tout d'abord pour appréhender les données avec leur spécificité et leurs biais éventuels, mais aussi pour comprendre le besoin opérationnel et/ou le contexte économique, social et politique.

Outre les différences mentionnées ci-dessus, la différence la plus importante entre le statisticien et le *data scientist* réside dans la finalité de leur action. L'un se place en observateur qui produit de la connaissance tandis que l'autre est focalisé sur le résultat en termes d'action et de conséquence opérationnelle de son travail. Ce trait majeur, même si les techniques appliquées sont proches, influe ensuite sur leur environnement, sur la communication de leur résultat et sur bien d'autres aspects de leur métier.

2.2.1.3 Des métiers de gouvernance

Il convient de souligner l'existence d'autres métiers en lien avec les données, notamment autour de la gouvernance des données. Le **chief data officer** est responsable des enjeux de gouvernance des données, de leur conservation, de leur dimension stratégique. Il endosse parfois un rôle technique en pilotant les choix notamment en matière d'architecture. Cette fonction inclut parfois une dimension d'innovation : pilotage et développement de prototypes, de preuves de concepts et de cas d'usages. Au sein de l'État, le rôle d'administrateur général des données (devenu AGDAC), d'administrateurs ministériels des données et d'administrateurs directionnels des données peuvent s'y apparenter.

Enfin, on note l'apparition depuis peu d'un nouveau métier, celui de **data steward**. Cet « intendant des données » peut avoir des rôles légèrement différents en fonction des organisations. De manière générale, il a en charge la qualité des données en assurant différentes tâches de suppression de doublons, de mise à jour, de contrôle qualité, de jointure, etc. Si ces tâches sont relativement proches de celles réalisées par le *data engineer*, le *data steward* offre un service qui n'est pas uniquement technique : il assure la documentation des données, aide à leur utilisation et à leur compréhension par les utilisateurs.

2.3 Quelles sont les spécificités de ces métiers ?

2.3.1 Les experts des données, des métiers numériques parmi d'autres ?

2.3.1.1 Une indéniable dimension numérique

La massification des données accompagne le développement du numérique. C'est donc en toute logique que les métiers d'experts de la donnée sont assimilés à des métiers du numérique (ce qui n'est pourtant pas le cas du métier de statisticien). *Data architects*, *data scientists*, *data engineers* et *data analysts* travaillent dans des environnements liés aux technologies de l'information et de la communication.

Le Répertoire Interministériel des Métiers de l'État (Rime) recense deux métiers spécifiquement liés aux données : le *data scientist* et le chargé de gouvernance de la donnée. Les deux sont classés dans la catégorie « numérique et systèmes d'information et de communication ». La circulaire du 19 mai 2017 portant sur le référentiel de rémunération des quinze métiers en tension de la filière numérique et des systèmes d'information et de communication intègre le métier de *data scientist*, le considérant donc également comme un métier du numérique.

L'inclusion des métiers d'experts de la donnée au sein des métiers du numérique se traduit parfois par une exploitation très large des compétences, en particulier dans le cas du *data scientist*. En effet, si l'objet et la production des autres métiers de la donnée (*data engineer*, *data architect*...) sont assez ciblés et correspondent en général à des recrutements de services ayant une expérience avancée en matière de donnée, le *data scientist* est parfois moins employé comme expert de la donnée que comme un expert du numérique et de l'innovation. Il lui est ainsi demandé d'accomplir, en sus de ses tâches premières de

traitement des données, d'autres tâches comme la mise en place d'API, d'outils d'ingestion de donnée et même parfois du développement d'applications (tâches qui relèvent davantage du *data engineer*, voire du développeur informatique). Cet usage extensif de la « ressource » *data scientist* n'est pas forcément pour déplaire aux individus, et donne plutôt satisfaction au service qui les emploie. Elle correspond en général à des services en cours de développement, qui mettent l'accent sur l'innovation, sans avoir toujours la possibilité de recruter l'ensemble des spécialistes.

Ce phénomène peut néanmoins traduire parfois une méconnaissance, par l'employeur, de son besoin et des différentes compétences qu'il doit réunir. De plus, la réalisation et le maintien dans le temps de travaux techniques réalisés par une personne qui n'est pas nécessairement experte et qui doit se former (ou, le plus souvent, s'auto-former) n'est pas un gage de qualité pérenne.

Par ailleurs, on peut parfois assister à un « déclasserment » des compétences dans lesquels le *data scientist* est mobilisé sur des tâches, telle la constitution de tableaux de bord simples ou l'extraction élémentaire de données, qui pourraient être réalisées par une personne moins qualifiée.

2.3.1.2 ... mais avec des spécificités

Les métiers d'experts de la donnée ont-ils des spécificités ou bien doivent-ils être considérés de la même manière que les *scrum master*, les responsables sécurités des SI ou les concepteurs-développeurs, autant de métiers qui apparaissent également parmi la liste des quinze métiers en tension ?

Le parallèle entre ces différents métiers est indiscutable lorsque l'on considère leur situation vis-à-vis du marché du travail. Il s'agit pour la plupart de métiers récents, en pénurie de main d'œuvre. Cela justifie en soi le traitement simultané de ces métiers lorsque l'on construit un référentiel de rémunération par exemple. La nécessité – et parfois la difficulté – d'évaluer les compétences techniques d'un profil candidat est un autre point commun aux métiers du numérique. Le nombre de pré-requis pour mener à bien son travail et la nécessité d'évoluer dans un environnement où il existe une culture et une compréhension de son métier peut également les rapprocher d'autres métiers du numérique.

En revanche, la mission considère que des différences existent pour le métier de *data scientist* et dans une moindre mesure pour celui de *data analyst* – d'ailleurs, d'après **l'enquête menée par la mission, la majorité des *data scientists* n'exercent pas dans les directions du numérique**. La dépendance de leur métier au nombre et à la qualité des données – dont ils doivent comprendre la sémantique - les distingue d'autres métiers moins dépendants d'une « matière première ». La présence dans l'administration d'un métier très proche, celui de statisticien, qui n'est pas dans la filière numérique, constitue une autre particularité. **Enfin et surtout, la difficulté d'évaluer la production du *data scientist* (cf. 2.5.3) constitue, comme pour le statisticien, une spécificité majeure de ces métiers.**

2.3.2 Des prérequis indispensables mais qui peuvent faire défaut

Les métiers de la donnée nécessitent pour leur exercice un certain nombre de pré-requis. Le *data scientist* a tout d'abord besoin de données. Pour lui comme pour le statisticien ou le *data analyst*, la donnée est la matière première qu'il va transformer et analyser. Le *data engineer* et le *data architect* interviennent plus en amont de la chaîne de la donnée mais ont, eux aussi, besoin d'infrastructures informatiques performantes.

Enfin, les experts de la donnée ne peuvent s'épanouir que dans un environnement professionnel et managérial qui dispose d'une compréhension ne serait-ce que minimale des enjeux des données et des moyens requis.

2.3.2.1 Les données

Dans certains cas, l'administration dispose de données de très bonne qualité permettant une mise en œuvre rapide de projets d'exploitation de données⁸. Cette situation – idéale – n'est cependant pas la plus courante dans le secteur public. Certains services de l'administration, qui avaient dans un premier temps recruté des

⁸ L'accès à des données d'une grande richesse peut même constituer un facteur d'attractivité pour des candidats potentiels.

data scientist de haut niveau, ont été contraints de changer d'approche, faute de données disponibles ou par manque d'investissement dans les infrastructures.

De nombreux freins à l'utilisation des données ont été rapportés à la mission parfois avec fatalisme : l'utilisation de référentiels non partagés, l'accès compliqué aux données détenues par d'autres administrations, l'absence de cartographie des données permettant d'enrichir son sujet, la non-disponibilité de standards de données, des données de qualité non homogène et sans documentation, etc. Ces éléments sont pourtant fondamentaux pour l'interopérabilité des données et, *in fine*, pour des travaux de qualité en *data science*. En effet, pouvoir assembler des données entre elles et recueillir des indices (appelées aussi co-variables ou *features*) supplémentaires constitue une grande partie de la valeur ajoutée des algorithmes de prédiction, classification, etc.

Ces éléments, déjà évoqués dans le rapport de la mission Bothorel, diminuent l'efficacité du travail des experts du traitement de données et engendre parfois des frustrations tant de la part de ces profils experts que des administrations qui les recrutent.

2.3.2.2 Les infrastructures et l'environnement de travail

Le traitement des données massives est une discipline relativement récente. Les besoins informatiques nécessaires ont plusieurs fois été rapportés lors des auditions comme n'étant pas au niveau. Les besoins des experts de la donnée dépasse l'informatique bureautique suffisante à la plupart des métiers de l'administration. L'outil de travail fourni aux experts de la donnée ne leur est pas toujours spécifique et ne présente pas toujours les caractéristiques nécessaires (mémoire, mémoire vive, l'accès au CPU) à leurs travaux.

De plus, même si l'expert en donnée a en général une meilleure connaissance de l'outil informatique, il n'a pas toujours plus de liberté sur son poste de travail que les autres métiers. Il a pourtant besoin d'installer des logiciels (R, Python et leurs librairies, Spark, Elastic Search, Dataiku, etc.) et de les paramétrer. La logique traditionnelle au sein de l'administration selon laquelle le service informatique est le mieux placé pour définir la qualité et l'intérêt d'un logiciel n'est pas la plus appropriée dans un contexte de logiciels variés nécessitant d'être testés et évalués en conditions réelles. Une partie de ces limitations est la conséquence – compréhensible et indiscutable – de considérations de sécurité informatique. Cependant, la spécificité des métiers de la donnée peut conduire à un traitement plus individualisé des situations.

Peut-être parce qu'ils ne sont pas toujours dans les directions du numérique, une partie non négligeable du temps des *data scientists* et de leur manager de proximité est dédiée à ces questions et au contournement des difficultés. Il s'ensuit parfois le sentiment de travailler au sein d'un environnement peu accueillant. Même s'ils ont conscience du risque que cela représente en matière de sécurité informatique, les agents concernés en viennent parfois à utiliser leur matériel personnel.

Enfin, certaines infrastructures de données peuvent parfois être considérées comme n'étant plus à l'état de l'art par les candidats aux postes de *data scientists*. Si le fait de travailler avec des bases de données classiques peut pénaliser le *data scientist* dans son travail, c'est surtout la peur de perdre en compétence en ne travaillant pas sur les nouvelles formes d'organisations de données (*Hadoop*, *Spark*) qui est soulignée par les candidats.

2.3.2.3 L'acculturation des administrations à la donnée

Le manque de culture de la donnée au sein des administrations est l'un des freins à la mise en œuvre des *data science* au sein de l'État.

« *L'utilité et le potentiel des données et des codes sont mal compris, et restent identifiés dans l'imaginaire collectif comme des sujets purement techniques, qui ne sont pas au service de l'élaboration et du pilotage des politiques publiques* » constatent les auteurs du rapport Bothorel, qui appellent au développement d'une politique de formation de la fonction publique plus ambitieuse sur les enjeux du numérique (obligation de

formation des cadres dirigeants aux enjeux du numérique, séminaires de cadres dirigeants, offre de formation pour tous les niveaux hiérarchiques, plans de formation ministériels complémentaires à l'offre interministérielle, modules dans l'ensemble des cursus de formation de la fonction publique)⁹.

Il convient de noter que le plan d'actions DGAFP-DINUM de 2019 comprend une action qui cible les cadres dirigeants (mise en œuvre d'un dispositif d'accompagnement des cadres dirigeants leur permettant d'intégrer l'impact du numérique sur la sphère publique). Cependant, il serait utile que ces actions soient élargies à l'ensemble des niveaux hiérarchiques. Dans le cadre du plan de relance, un programme budgétaire est d'ailleurs dédié aux actions d'accompagnement des agents dans l'appropriation des enjeux du numérique. Cependant, force est de constater que, parmi la première vague des projets retenus, un seul concerne spécifiquement les enjeux de la donnée¹⁰.

Le développement d'une culture de la donnée ne doit pas viser à former des profils d'experts (*data scientists*, *data engineers*), mais plutôt à faire en sorte que l'ensemble des agents, en commençant par les cadres dirigeants, prennent conscience du potentiel des données dans le cadre de leurs missions. L'IGPDE a d'ailleurs entrepris le développement d'actions de formation et de sensibilisation en ce sens (notamment le cycle supérieur du numérique), et la stratégie de l'Intelligence artificielle au service de la Défense engage également la volonté de diffuser une culture volontariste d'usage de l'IA dans le ministère en confiant à la Cellule de Coordination de l'Intelligence Artificielle de Défense (CCIAD) la mission de sensibiliser le management de haut niveau du Ministère des Armées.

Afin de permettre un développement pertinent de l'usage des données, la mission ne peut qu'encourager les initiatives visant une meilleure acculturation à la donnée. Si une partie des formations pourrait s'appuyer sur un programme commun avec une orchestration au niveau interministériel, la spécificité de chaque ministère avec ses organismes de formation et son AMD milite pour un pilotage au niveau ministériel. Les feuilles de route ministérielles qui doivent être remises au Premier ministre et à la Ministre de la transformation et de la fonction publique au 15 juillet et rendues publiques le 15 septembre 2021 contiendront très probablement une action en ce sens. La mission recommande un objectif chiffré en termes de formation :

Recommandation 1 : Former à la donnée 50 % des cadres dirigeants d'ici fin 2022, 90 % d'ici 2024.

Ces formations devront contenir des modules sur l'usage des données, sur les technologies, présenter des retours d'expérience de projets menés au sein des administrations, inclure une sensibilisation aux enjeux juridiques et éthiques ainsi que des éléments sur la réalisation de projets mettant en œuvre des technologies d'intelligence artificielle.

2.4 Des exigences particulières et des risques à prévenir

Mener des projets de *data science* nécessite non seulement des compétences expertes et un environnement technologique adapté, mais aussi la capacité à prévenir un certain nombre de risques susceptibles de compromettre la réussite des projets ou la pertinence des résultats obtenus.

La puissance des outils de traitement des données ne suffit pas à assurer la pertinence des résultats, Il faut aussi maîtriser le contexte dans lequel s'inscrivent les données traitées, apprécier leur représentativité et savoir interpréter les résultats obtenus. Un cadre rigoureux et maîtrisé d'exercice des métiers de la donnée est à ce titre nécessaire. De même que le pilotage des sous-traitants exige de la part des maîtres d'ouvrage une expertise et un recul suffisants.

⁹ Recommandation n°17, rapport « Pour une politique publique de la donnée » dit rapport Bothorel

¹⁰ Formation au langage R pour les économistes du Trésor, projet financé dans le cadre du sous-programme ITN-5 du plan de relance – cf. chapitre 1

2.4.1 La nécessité de comprendre et d'adapter les méthodes au problème posé

Dans l'évaluation de leur travail, on entend parfois des *data scientists* se satisfaire de la mise en place d'un modèle (réseau de neurones, arbres aléatoires, xgboost, etc.) qui prédit bien telle ou telle métrique avec un taux de succès de 96 %. De leur point de vue, ils ont rempli la mission qui leur était confiée, à savoir de prédire au mieux telle variable avec telle donnée.

Pour autant, le choix de la bonne métrique est crucial. Elle doit être celle qui correspond au problème posé et non celle que l'on maîtrise le mieux ou la plus facile à en mettre en place et déjà programmée. **L'encadrement a un rôle important à jouer** pour expliciter les objectifs, de même qu'est essentiel le dialogue avec les directions métiers utilisatrices des résultats.

Si on veut prédire à partir d'un historique de données quel camion de pompier pourrait arriver le premier sur un incident il faut au préalable déterminer le risque que l'on veut maîtriser : le temps total d'arrivée sur les lieux ? Mais est-ce qu'on ne préfère pas envoyer des camions qui arrivent un peu plus lentement en moyenne mais jamais en plus de vingt minutes ? Si tel est le cas, il va falloir adapter les méthodes en conséquence et proposer une solution qui s'écarte du cadre théorique général¹¹.

Par ailleurs, l'apprentissage sur des données passées pour établir un service présent suppose une continuité dans le temps du service or, bien souvent, on souhaite une rupture avec le passé parce que l'algorithme doit permettre des décisions plus objectives ou bien parce que l'algorithme est établi dans le cadre de la modification d'un processus ou d'une nouvelle politique publique. Il faut alors prendre en compte le fait que ce changement aura une incidence sur les données et qu'il n'est peut-être pas possible d'apprendre immédiatement. Il faut alors pouvoir dire que le projet n'est pas réalisable dans l'immédiat ou savoir anticiper la réaction au changement en exploitant parfois des modèles comportementaux et ne pas être dans le cadre du seul apprentissage automatique sur les données¹².

Le **data scientist doit donc maîtriser les risques liés à son activité**¹³. Il évolue dans un contexte particulier car certains algorithmes d'apprentissage fonctionnent comme des « boîtes noires » et il n'existe pas de théorème mathématique qui permette d'assurer leur validité ou d'évaluer leur incertitude. Le *data scientist* doit alors identifier les variables qui participent à la décision de l'algorithme, pour en comprendre le fonctionnement et ainsi convaincre le métier et se convaincre lui-même que la « boîte noire » est fonctionnelle.

2.4.2 Les risques à ne pas appréhender les biais des données

Une littérature émergente fait état des biais, des erreurs récurrentes et des exemples d'échec liés à l'intelligence artificielle. Certains auteurs, comme Cathy O'Neil, tentent de faire comprendre à un large public les risques liés à l'intelligence artificielle. Les biais sont généralement liés à la qualité des données.

On peut penser à des biais involontaires lorsque les données d'apprentissage ne sont pas représentatives (seulement un sous-ensemble de la population cible a généré les données) ou en cas de recueil imparfait des données d'apprentissage (décalage entre la réalité et celle captée par les données). Par exemple, l'ouvrage mentionné ci-dessus évoque un algorithme de police prédictive aux États-Unis qui se fonde sur les crimes qui ont été recensés dans une zone géographique donnée. Parfois, la majorité de la base de données

¹¹ Les algorithmes d'apprentissage automatique ont en général une métrique d'erreur simple et n'intègre pas de contrainte.

¹² Dans ce contexte, le positionnement des services statistiques peut être un atout. En effet, ils ont l'expérience – et le devoir – de parfois refuser de communiquer des productions lorsque la précision des données est trop faible ou les biais trop importants. Même si le contexte est différent, il peut exister une pression sur les *data scientists* pour produire des algorithmes en dépit des risques et avoir parmi eux des personnes ayant une certaine forme d'indépendance est un élément rassurant.

¹³ Un courant de recherche étudie les échecs de l'intelligence artificielle pour déterminer les moyens de les prévenir. Une base de donnée sur le sujet a d'ailleurs été constituée. Il y a un consensus pour classer les échecs en deux catégories : ceux liés à des erreurs, déjà évoquées venant d'un décalage entre les données d'apprentissage et le contexte d'utilisation, de la réalité non optimale captée par ces données et des configurations qui n'ont pas été rencontrées jusqu'à présent. Mais aussi les échecs liés à des attaques (on « empoisonne » les données d'apprentissage pour manipuler, on cherche à comprendre le modèle pour pouvoir en jouer par exemple), etc.

est constituée par des infractions mineures ce qui peut conduire les crimes plus importants, ceux que la police cible en priorité, à être « perdus dans la masse » et finalement non traités par l'algorithme.

Le thème de la police prédictive est aussi remarquable pour expliquer le biais d'endogénéité, connu des économètres ; la présence ou non d'un crime dans une base de données est liée d'une part au crime mais aussi à la présence policière à la fois dans ce qu'elle dissuade et dans ce qu'elle permet de constater les infractions. Le ciblage par la police d'un quartier donné aura donc une influence sur les données de ce quartier : un algorithme d'apprentissage mal calibré – celui qui ne prend pas en compte cet effet – conduira nécessairement à une situation erronée. Ou bien, il conseillera aux policiers de rester indéfiniment dans le quartier où ils constatent les infractions et, ce faisant, renforcer son idée que c'est dans ce quartier qu'il y a des infractions et que les policiers doivent aller. Ou bien, à l'inverse, si l'effet dissuasif de la police domine, l'algorithme constatera que la criminalité augmente dans les quartiers où la police n'est pas, et suggérera simplement aux forces de polices d'aller là où elles n'étaient pas dans la période précédente sans prendre en compte la hausse de la criminalité qui est susceptible d'advenir dans le quartier délaissé suite à sa recommandation. Dans les deux cas, l'impact sur la criminalité sera réduit.

Enfin, l'apprentissage automatique, par construction, reproduit les éléments du passé et peut donc prolonger les biais de sa base d'apprentissage. S'il y a un biais raciste dans les délits rapportés dans les bases de données, spontanément l'algorithme va reproduire ce biais. Cela rejoint le fait que la plupart du temps, les données sont contextualisées – attachées à un contexte particulier – et qu'il faut **maîtriser le contexte pour rectifier la base de données et l'apprentissage**.

2.4.3 Un enjeu d'équité incontournable

Dernier enjeu de taille : avec l'augmentation de la complexité des algorithmes, il devient difficile de leur faire intégrer des principes d'équité pourtant essentiels, a fortiori pour le service public. Se développent alors une nouvelle thématique de recherche : l'explicabilité de l'algorithme qui permet de lui faire confiance. Sam Corbett-Davies et Sharad Goel proposent dans « The Measure and Mismeasure of Fairness : A Critical Review of Fair Machine Learning », une revue de littérature.

Du point de vue de l'ingénieur, l'équité de l'algorithme d'apprentissage est une contrainte pour l'optimisation et il s'agit donc de sacrifier un peu de la précision du modèle pour asseoir cette équité. Par ailleurs, la notion d'équité est subjective, elle doit être déterminée collectivement et non pas par le *data scientist* seul. Cela d'autant plus que toutes les formes d'équité ne peuvent être assurées simultanément. Par exemple, un algorithme de sélection de candidats à une offre d'emploi ne peut pas à la fois garantir qu'autant de travailleurs handicapés auront un emploi que ceux sans reconnaissance de handicap et que la probabilité de recruter l'un ou l'autre soit la même.

Même si elles demeurent relativement confidentielles, les initiatives de « serment d'Hippocrate » du *data scientist*, font un lien intéressant entre la responsabilité de l'expert en intelligence artificielle et le médecin dont le pouvoir doit être accompagné d'un sens éthique fort.

2.4.4 L'isolement du *data scientist* est un facteur de risque

La vision du *data scientist*, simple ingénieur mathématicien face à un problème bien déterminé, est loin de la réalité du métier. Si la maîtrise de la technique est une condition nécessaire, elle n'est pas suffisante. Le *data scientist* doit, comme en statistique, **exercer son métier sous la responsabilité d'experts** qui ont eu cette technicité et qui, avec l'expérience, sont en mesure de prendre du recul. Avec ce recul, il est alors possible d'encadrer le travail, de traduire des orientations opérationnelles, éthiques et politiques en termes techniques mais aussi de traduire les choix techniques qui doivent être faits auprès des autres membres de l'organisation.

Le partage d'expérience, le dialogue entre pairs est aussi important dans un domaine fortement évolutif et pour lequel le *data scientist* n'est pas nécessairement en mesure de maîtriser l'ensemble des techniques avec la même acuité.

2.5 Une compétence difficile à évaluer

2.5.1 Au moment du recrutement

L'émergence d'un nouveau métier technique pose la question du recrutement de spécialistes par des personnes qui, par construction, ne maîtrisent pas la dite technique. Si le secteur privé fait parfois appel à des cabinets de recrutements spécialisés et sous-traite une partie de l'évaluation des candidats, cette solution n'est guère utilisée sous cette forme dans l'administration publique. Cependant, il a été rapporté à la mission que, plusieurs fois, le service statistique ministériel (SSM), l'administrateur ministériel des données (AMD) ou un autre service avait assuré une prestation de ce type pour une entité débutant un recrutement de spécialiste de la donnée.

Si les managers directs conservent la responsabilité de l'évaluation « humaine » du candidat, l'évaluation technique est alors réalisée à part, souvent à partir d'une mise en situation et d'un test technique. L'évaluation séparée des compétences techniques est la pratique la plus répandue dans les organisations publiques ou privées que la mission a pu interroger.

Cependant, le recrutement peut aussi être réalisé sans appui technique extérieur. Parfois, le recruteur, spécialiste métier ou spécialiste du numérique, n'a pas les compétences en données nécessaires à une évaluation technique des candidats *data scientists*. Dès lors, ce sont parfois les premières personnes recrutées au sein d'une équipe de *data science* qui se chargent d'évaluer leurs futurs collègues. Cette solution, qui peut laisser une place importante à la subjectivité, peut présenter certains risques.

2.5.2 Et sur les résultats

Il y a une différence majeure qui distingue le *data scientist* des autres métiers du numérique et de beaucoup de métiers : l'évaluer n'est pas aisé. Lorsque la production est un outil ou une application, on peut évaluer sa qualité en l'utilisant, lorsque la production est un nombre, pour le statisticien, ou un algorithme pour le *data scientist*, l'évaluation n'est pas intuitive. Certes, dans sa manière de travailler, le *data scientist* utilise une partie des données pour évaluer son algorithme de prédiction mais il est difficile de savoir précisément ce qu'il en sera dans le temps, après la mise en production. L'évaluation par une métrique calculée sur les données suppose que cette métrique soit judicieusement choisie, or cela requiert une expertise certaine.

D'autre part, celle-ci est parfois trop simple par rapport à la réalité et ne prend pas en compte tous les paramètres. En 2014, par exemple, la société Amazon a mis en place un projet de tri des CV pour faciliter ses recrutements. Entraîné sur les recrutements passés de la société, l'algorithme exploitant le texte des offres a interprété à tort une corrélation entre le genre du candidat et sa probabilité d'être retenu avec pour conséquence de systématiquement déclasser les candidatures féminines pour certains postes.

La fiabilité dans les programmes d'apprentissage automatique est aujourd'hui un enjeu de recherche et il n'est pas toujours possible de la garantir. Comme le statisticien doit parfois expliquer qu'une donnée n'est pas fiable et qu'on ne peut construire un raisonnement ou une politique publique sur son évolution, **le *data scientist* devra pouvoir dire que le projet n'est pas réalisable**. Il ne s'agit en effet pas seulement de développer un algorithme mais d'évaluer sa validité dans le temps et d'anticiper des changements de comportements, des changements de politiques publiques ou des changements de systèmes d'information qui peuvent altérer la pertinence de l'algorithme établi. Entamer des recherches, investir, et finalement considérer que le projet n'est pas viable nécessite de pouvoir être en situation de le dire et de ne pas être pénalisé, au contraire, par cette issue.

À mesure que les projets sur les données vont monter en puissance, ces sujets d'évaluation vont devenir de plus en plus importants. Comment garantir la viabilité d'un algorithme qui fait un tri sur les patients dans les services d'urgences ? Qui affecte une caserne et un véhicule de secours à une urgence ? Qui fait de la maintenance prédictive sur des rails ? Dans ces cas, comme dans bien d'autres, l'erreur d'un algorithme ou son biais peut avoir des conséquences tragiques.

Ces éléments montrent, en particulier pour le *data scientist*, l'importance de disposer de compétences de haut niveau au sein de la fonction publique pour répondre aux enjeux de qualité mais aussi éthiques posés par les travaux de science des données. Le chapitre 3 vise à quantifier le besoin de compétences et le chapitre 4 explore et recommande des pistes pour assurer des compétences en nombre et qualité au sein de la fonction publique.

Chapitre 3 : Répondre à la croissance du besoin de compétences

3.1 Une nette montée en puissance depuis 2017

3.1.1 Des projets qui se multiplient

On assiste depuis 2018 à une nette montée en puissance des projets de transformation par la donnée dans les administrations de l'État, comme en atteste le succès des différents dispositifs de soutien et de financement ouverts :

- Avec le développement du dispositif des Entrepreneurs d'intérêt général depuis 2017, les projets accompagnés de *data science* et data ingénierie se sont multipliés. Sur un total de 70 défis, ce sont à ce jour 32 projets dans ces domaines qui ont bénéficié du dispositif, dont 25 pour des services de l'État ;
- 11 des 21 projets lauréats des appels à manifestation d'intérêt IA en 2019-2020 et bénéficiant à ce titre d'un accompagnement d'Etalab, concernent des administrations de l'État, les autres projets étant principalement portés par des opérateurs de l'État ;
- S'agissant des projets financés par le Fonds de transformation de l'action publique (FTAP), une proportion très significative recouvre à titre principal ou secondaire un enjeu autour des données : projets « big data et IA », création d'infrastructures de données, appariements de données, mise à disposition de données... Parmi les projets les plus ambitieux qui ont bénéficié du fonds, dix projets d'administrations de l'État¹⁴ représentent par exemple à eux seuls un financement de plus de 100 millions d'euros du FTAP, ces projets étant portés par la DGFIP, la DGDDI, la DGCCRF, la Direction des affaires maritimes, la Direction du numérique du Ministère de l'Intérieur et la Drees pour le Health Data Hub ;
- Dans le cadre du volet « développer l'utilisation de la donnée dans votre administration » du plan de relance, une vingtaine de projets portés par des services de l'État ont d'ores et déjà été retenus.

Ces projets présentent une grande diversité à la fois dans leur contenu et du point de vue des services qui les portent, la plupart des ministères étant concernés. Ils témoignent de l'appropriation grandissante au sein de l'État des enjeux autour de la donnée, avec un développement de l'identification de cas d'usage pertinents.

3.1.2 Des recrutements en nette augmentation au cours des dernières années

La mission a cherché à apprécier l'évolution des emplois d'experts de la donnée en étudiant les offres d'emploi publiées sur la Place de l'Emploi Public depuis 2016. Cette approche permet de fournir des éléments historiques et de présenter une tendance que l'enquête réalisée par la mission auprès des administrations complète par des éléments précis sur le présent et des projections sur l'avenir.

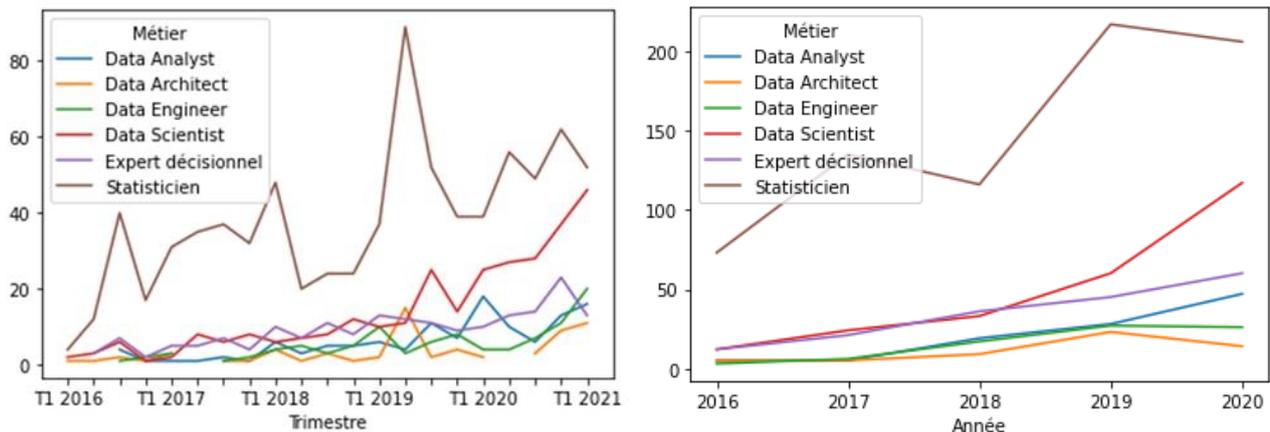
L'enquête est indispensable car l'approche via les offres d'emploi n'est pas exempte de faiblesse méthodologique, la principale étant que l'on comptabilise les recrutements là où la mission s'intéresse aux postes. Il est également possible que certains postes, peut-être davantage ceux réservés aux fonctionnaires, ne donnent pas lieu à une publication sur la place de l'emploi public. Certains postes qui ne sont pas ouverts aux contractuels peuvent donc ne pas être comptabilisés, quand d'autres postes ayant connu une forte rotation sont comptabilisés plusieurs fois, à chaque fois que le poste est vacant.

¹⁴ 8 projets de la catégorie « Big data et IA » portés par des administrations de l'État, le projet 3D de l'administration des Douanes et le projet Pilat de la DGFIP

Compte tenu des limites de l'approche, la mission considère qu'il est plus pertinent d'observer l'évolution des recrutements plus que leur niveau en valeur absolue.

Le graphique 1 présente les résultats pour les catégories de métiers définis par la mission mais y ajoute une courbe spécifique aux experts décisionnels, appelés parfois experts en *business intelligence*. Il est en effet difficile de distinguer dans les offres la composante de mise en place du système décisionnel (qui relève de l'ingénierie de la donnée) et celle de la réalisation de tableaux de bord à partir du système déployé (plus proche du métier de *data analyst*).

Graphique 1 : Offres d'emploi par métier au sein de l'État par trimestre entre 2016 et le premier trimestre 2021 et par année entre 2016 et 2020.



Source : Place de l'Emploi Public, calculs Mission (voir annexe 4 pour plus de précision).

Le nombre d'offres d'emploi est en forte croissance depuis 2018 pour les métiers d'experts de la donnée. Le nombre d'offres d'emploi pour les *data scientists* est passé de 33 en 2018 à 117 en 2020, pour les *data analysts* le nombre a augmenté de 5 en 2017 à 47 en 2020. Le nombre d'offres pour des postes de *data engineers* progresse lentement pour atteindre 26 offres en 2020 (moins de dix recrutements en 2016 et 2017) tandis que le nombre d'offres relatives aux *data architects* n'est plus tout à fait confidentiel passant de 5 en 2016 à 15 en 2020. Les offres correspondant à des experts décisionnels sont en progression constante entre 2016 et 2020. Au total, on recense pour l'année 2020 267 offres sur ce métier, contre 37 offres seulement en 2016.

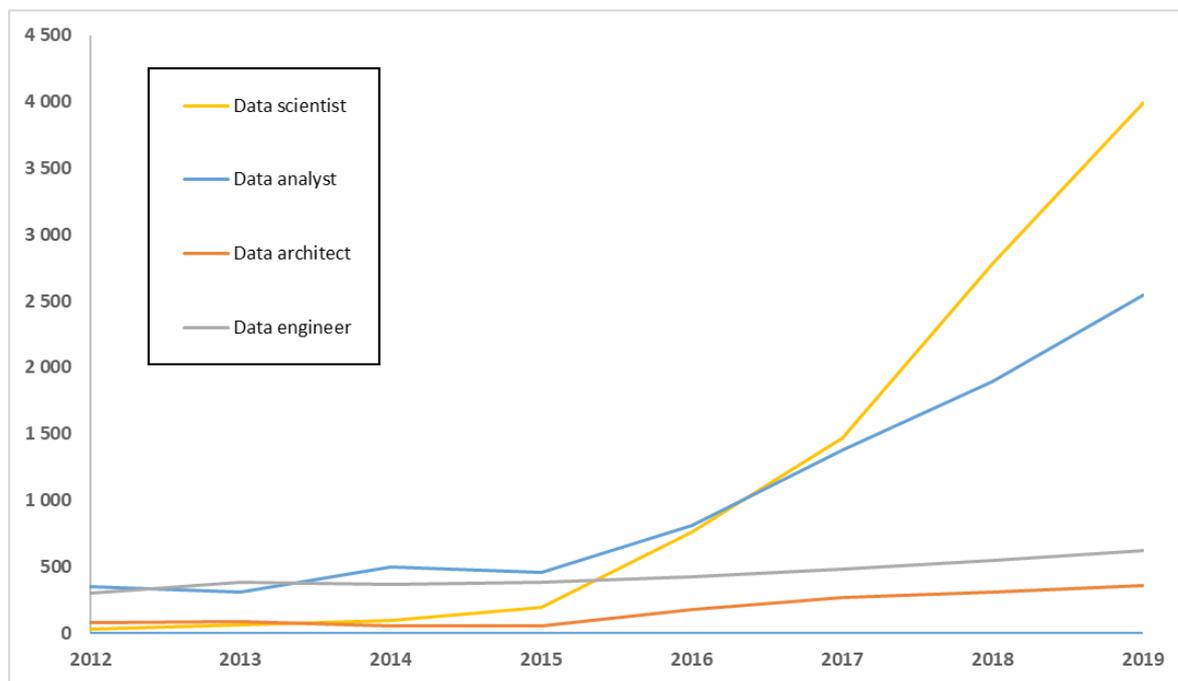
La répartition des offres par ministère montre une concentration sur quelques domaines ministériels. Plus de 50 % des offres se concentrent uniquement dans le champ de trois ministères : Ministère de l'Éducation Nationale de la Jeunesse et des Sports (regroupé dans les données avec le Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation) ; le Ministère de l'Intérieur et les Ministères Économiques et Financiers. Les offres concernent néanmoins un spectre très large de ministères et au sein de chacun d'entre eux, une variété de directions.

Même si le nombre d'offres correspondant au métier de statisticien est particulièrement mal appréhendé (cf. annexe méthodologique), tous les calculs réalisés montrent que, dans le même temps, le nombre d'offres correspondantes croît également sur la période 2016-2020.

Peut-on mettre en perspective cette évolution ? La mission a exploité les données du recensement de la population en reprenant la méthodologie de l'article *Data scientists, community managers... et informaticiens : quels sont les métiers du numérique ?*¹⁵ réalisé par la Dares en 2019.

¹⁵ Data scientists, community managers... et informaticiens : quels sont les métiers du numérique ?, Aurore Desjonquères, Claire de Maricourt, Christophe Michel, *L'économie et la société à l'ère du numérique*, édition 2019 ; <https://www.insee.fr/fr/statistiques/4126588>

Graphique 2 : Effectif par métier dans la population entre 2012 et 2019



Source : Enquêtes Annuelles du recensement (Insee), traitement de la mission

La hausse du nombre d'offres diffusées sur la Place de l'Emploi Public depuis 2018 semble suivre un mouvement débuté en 2016 dans la sphère privée. Si la source de données permet en théorie d'identifier la part de l'emploi public dans les métiers concernés, les effectifs (moins d'une centaine) sont trop faibles pour pouvoir être bien appréhendés. Les statisticiens représentent cependant une exception. Non présenté dans le graphique ci-dessus par souci de lisibilité, le nombre de statisticiens est constant ces dix dernières années et avoisine les 7 000 personnes parmi lesquelles un peu plus du quart travaillent dans le secteur public.

3.2 Une enquête qui permet pour la première fois de dresser un tableau des compétences « data » au sein de l'État

La mission a adressé un questionnaire à tous les administrateurs ministériels des données pour recenser la situation actuelle et les besoins de compétences en expertise de la donnée (cf. en annexe 5 le questionnaire). Il est apparu que ceux-ci ne disposaient pas toujours de l'information et ont eu parfois du mal à transmettre une réponse. Cependant, la quasi-totalité des ministères et les principales directions sollicitées ont répondu, montrant le fort intérêt pour ce sujet. L'enquête concernait l'ensemble des effectifs des ministères en administration centrale et dans les échelons déconcentrés. Néanmoins, les réponses sont concentrées sur l'administration centrale pour deux raisons : premièrement il était plus difficile d'obtenir l'information sur les échelons déconcentrés dans le temps imparti, deuxièmement il y a probablement relativement peu d'experts de la donnée en dehors de l'administration centrale sauf exceptions. Parmi les directions les plus concernées, l'enquête tient compte des experts de la donnée de la DGFIP situés en régions mais pas ceux des DREETS (Directions régionales de l'économie, de l'emploi, du travail et des solidarités).

Pour compléter le panorama, un questionnaire identique a été envoyé aux entités concernées *a priori* par le sujet et ne dépendant pas d'un ministère, ou pour lesquelles un AMD n'a pas été identifié (services du Premier ministre et autorités administratives indépendantes).

En outre, un questionnaire a été adressé à l'Insee et aux services statistiques ministériels (SSM). Il comportait à la fois des questions similaires de façon à pouvoir compiler les réponses et des questions spécifiques notamment sur le positionnement des SSM. Une grande partie des SSM y a répondu. Le champ couvre les échelons centraux et une partie des échelons déconcentrés : ceux du SSM de l'agriculture et des pôles nationaux de l'Insee situés en régions. En revanche, il ne couvre pas le reste des effectifs des directions régionales de l'Insee, notamment les services d'études et les services de production de données. Ceci explique que l'enquête ne retrouve pas tous les cadres A des corps de l'Insee (outre le fait qu'une partie des cadres A de ces corps n'est pas sur des postes d'experts de la donnée).

Au total, les réponses couvrent une très grande partie des services de l'État (cf. annexe 6 liste des répondants). Après compilation des réponses et traitement des incohérences, la mission considère que, malgré une couverture non exhaustive et des définitions de métiers non stabilisées, **l'enquête fournit des ordres de grandeur qui permettent de dresser un tableau des compétences data au sein de l'État.**

3.2.1 Des besoins fonctionnels couvrant un large spectre

Les réponses témoignent que de nombreuses directions ont besoin de compétences en donnée, lesquelles servent donc un très large spectre de besoins pour éclairer des décisions d'un grand nombre de métiers. Les besoins qui se retrouvent le plus souvent sont l'amélioration de la qualité des données, la production de tableaux de bord, la cartographie et la production de statistiques. Ceux qui nécessitent des compétences en intelligence artificielle touchent davantage à l'analyse prédictive, à l'optimisation des processus et à la détection de valeurs atypiques.

Par exemple, plusieurs dispositifs visent à

- mieux détecter les fraudes dans les déclarations, comme celles relatives aux revenus ou au chômage partiel, et cibler les contrôles ;
- construire un référentiel indicatif d'indemnisation de préjudices corporels ;
- mieux piloter et défendre un navire militaire ;
- modéliser des prédictions robustes intégrant des facteurs externes multiples dans la gestion de la crise du Covid.

3.2.2 Des profils très divers pour un effectif global de 2 000 experts

Pour répondre à ces besoins, au total, il y a au sein de l'État environ 2 000 personnes qui travaillent comme experts de la donnée d'après l'enquête conduite par la mission. Il y en a en réalité davantage puisqu'il manque les personnes non couvertes par l'enquête (comme les services de renseignement, voir annexe 6 la liste des répondants à l'enquête). Parmi les 2 000 recensés, **1 300 se trouvent au sein de l'Insee et des services statistiques ministériels (SSM) et environ 700 en dehors.** Ces derniers représentent une petite partie des 18 000 experts du numérique identifiés par la DINUM.

En raison de la relative nouveauté du sujet et de son évolution encore forte au fil des ans, il n'existe pas de référentiel interministériel de métiers exhaustif dans le domaine de la donnée. Cela rend difficile la segmentation et l'identification des différents métiers. En dehors de l'administration, il existe un certain nombre de référentiels, malheureusement non concordants.

La mission a choisi un découpage parmi d'autres, celui de l'Apec, en adaptant son contenu au secteur public, afin de compiler les résultats des différentes entités. Celui-ci comprend quatre métiers principaux ; *data scientist*, *data analyst*, *data engineer* et *data architect* (cf. le chapitre 2 pour leur définition) auxquels la mission a ajouté un cinquième métier dit polyvalent correspondant aux personnes à cheval sur l'analyse et l'ingénierie. Afin de correspondre à toutes les situations, un métier « autre expert de la data » a été proposé.

Malgré tout, le découpage utilisé pose problème quand il ne correspond pas à une grille ministérielle ou quand les ministères n'ont pas les mêmes notions des métiers. Il arrive même que la séparation entre deux métiers ne soit pas vue de la même façon par différents acteurs au sein d'un même ministère. La mission a procédé à quelques redressements pour résoudre les incohérences.

Il a tout de même été possible d'évaluer un ordre de grandeur des effectifs des différents métiers, afin de dégager des tendances. **En dehors de l'Insee et des SSM** (cf. tableau 1), **le métier le plus fréquent est celui de *data analyst* dont la moitié des effectifs exercent dans deux ministères** (le Ministère de l'Économie, des finances et de la Relance et le Ministère des Armées). Les auditions ont confirmé la prépondérance des *data analysts* en précisant qu'il existe des profils très variés allant des experts pointus aux *data analysts* moins experts mais dédiés à des travaux indispensables sur les données. **Le deuxième métier le plus fréquent est celui de *data scientist*** avec, là encore, plus de la moitié des effectifs dans les deux mêmes ministères. Les *data engineers* et les polyvalents sont moins nombreux, les *data architects* sont beaucoup plus rares. Les autres experts de la donnée occupent des fonctions très dispersées, la plus courante étant la tenue de répertoires.

Tableau 1 : Experts de la donnée par métier au sein de l'État hors Insee et SSM

	Effectifs en équivalent temps plein en mars 2021
<i>Data scientists</i>	170
<i>Data analysts</i>	240
<i>Data engineers</i>	100
<i>Data architects</i>	20
Polyvalents	70
Autres experts de la donnée	90
Total experts de la donnée	690

Source : réponses au questionnaire de la mission et estimations des auteurs

La comparaison avec une enquête précédente suggère un développement du nombre de *data scientists* et de *data engineers*. Le suivi du plan d'action national pour l'Intelligence artificielle en 2019 en dénombrait 85 en 2019 sur un champ différent de notre enquête, en termes de métiers – plus proches de l'IA – et en termes de ministères concernés – moins nombreux. Ils sont 280 en 2021 d'après notre enquête : cette augmentation reflète un élargissement du champ couvert par les enquêtes et une réelle augmentation, sans pouvoir distinguer les deux effets.

Pour l'Insee et les SSM, la distinction entre les métiers est un peu différente parce qu'ils emploient des statisticiens experts de la donnée en grand nombre et parce que ces derniers ont des compétences multiples en acquisition, traitement, analyse et valorisation de la donnée. En raison de leur formation, ils sont capables d'analyser les données dans un contexte économique et social. Les statisticiens sont partagés par l'enquête en deux groupes selon les compétences techniques mises en œuvre : ceux qui occupent des fonctions de *data scientists* et les autres qui sont regroupés avec les *data analysts*. Le partage n'est pas simple parce que nombre de statisticiens sont à cheval entre *data scientist*, *data analyst* et parfois *data engineer*. Une partie des statisticiens, comme les économètres ou des chargés d'études, conçoivent en outre des modélisations complexes même si elles ne sont pas considérées comme relevant de la *data science*.

La mission a reçu de multiples questions pour déterminer dans quel métier classer un statisticien selon l'activité qu'il menait et a partagé les réponses avec l'ensemble des services enquêtés. Cependant, plusieurs employeurs n'ont pas été en mesure dans leur réponse de classer les statisticiens experts de la donnée selon les différents métiers proposés.

Moyennant quelques hypothèses, on peut néanmoins considérer qu'une grande majorité d'entre eux occupent plutôt des fonctions se rapprochant davantage des métiers de *data analysts* que de *data scientists* (cf. tableau 2). C'est une vision métier mais une vision sur les compétences détenues – et pas seulement mises en œuvre – donnerait une autre image. En effet, les statisticiens sortis d'école ces dix dernières années ont reçu une formation pour faire de la *data science* à un niveau très pointu. Les statisticiens des

générations précédentes pourraient également en faire pour une bonne partie moyennant une formation de quelques semaines.

Tableau 2 : Experts de la donnée par métier au sein de l'Insee et des SSM

	Effectifs en équivalent temps plein en mars 2021
Statisticiens <i>data scientists</i>	120
Statisticiens et <i>data analysts</i>	910
<i>Data engineers</i>	120
<i>Data architects</i>	20
Polyvalents	70
Autres experts de la donnée	90
Total experts de la donnée	1 330

Source : réponses au questionnaire de la mission et estimations des auteurs (le champ Insee n'a pas été couvert entièrement)

Au total au sein de l'État, y compris l'Insee et les SSM, les *data analysts* (compte tenu des statisticiens du SSP non *data scientists*) sont les plus nombreux (1 150) devant les *data scientists* (290) et les *data engineers* (220).

3.2.3 Hors Insee et SSM, un recours légèrement minoritaire à des fonctionnaires

3.2.3.1 De forts contrastes dans les profils d'experts

Dans les SSM, les fonctionnaires constituent les deux tiers des effectifs et à l'Insee encore davantage en raison notamment des nombreuses personnes issues des corps de l'Insee. Les contractuels, aux deux tiers en contrat à durée déterminée, représentent un quart des experts de la donnée des SSM.

En dehors de l'Insee et des SSM, la situation est plus contrastée. Les fonctionnaires forment près de la moitié des effectifs, part qui atteint 60 % si on y ajoute les contrats à durée indéterminée (cf. tableau 3). Les corps les plus cités sont ceux de l'Insee, puis les ingénieurs des ponts, des eaux et des forêts, les attachés d'administration centrale, les ingénieurs des systèmes d'information et de communication. Les corps ministériels ou directionnels sont mécaniquement moins cités tels ceux des travaux publics, des douanes, de la DGFIP ou des officiers, même si certains emploient ou vont employer des experts de la donnée en nombre comme celui de la DGFIP. Les contractuels en CDD constituent un peu plus de 15 % des effectifs et les prestataires un peu moins. Les autres statuts internes ou externes (apprentis, entrepreneurs d'intérêt général, chercheurs extérieurs...) sont peu représentés.

Tableau 3 : Experts de la donnée par statut au sein de l'État hors Insee et SSM

en ETP en mars 2021	Effectifs	%
Fonctionnaires	320	46 %
CDI	90	13 %
CDD	110	16 %
Autres internes (apprentis, EIG...)	50	7 %
Prestataires	100	14 %
Autres externes (conventions...)	20	3 %
Total	690	100 %

Source : réponses au questionnaire de la mission et estimations des auteurs

3.2.3.2 Les *data scientists* beaucoup moins souvent fonctionnaires

Alors que près de la moitié des experts de la donnée sont fonctionnaires, ce n'est le cas que de moins d'un quart des *data scientists*. Les *data scientists* sont au contraire beaucoup plus souvent des contractuels en CDD, et principalement à la DGA en CDI, ou des ressources internes issues de dispositifs plus rares pour les autres métiers (apprentis, EIG...).

Les autres métiers sont exercés pour plus de la moitié par des fonctionnaires, à part la catégorie des *data engineers* constituée pour un tiers de fonctionnaires et pour un autre de prestataires. Cet important appel à des ressources externes est beaucoup plus fort que pour les autres métiers.

3.2.3.3 Une différence très significative entre les ministères

Il existe cependant une très importante différence d'approche en termes de statuts entre les ministères ayant une culture scientifique et des corps techniques, et les autres ministères. Seuls les premiers disposent d'un vivier interne qui peut fournir des experts de la donnée moyennant un investissement de formation pour les mettre à niveau. De fait, ils ont beaucoup plus d'experts de la donnée fonctionnaires que les autres. Par exemple, les fonctionnaires représentent plus de 80 % des experts de la donnée au Ministère de la Transition écologique contre moins de 10 % dans les ministères sociaux (hors SSM).

3.2.4 Des équipes qui se structurent de façon relativement diffuse

Il existe une grande dispersion des compétences *data* dans une multitude d'unités ainsi que plusieurs points de concentration situés à l'Insee et dans les services statistiques ministériels pour les plus importants (entre 50 et 150 experts de la donnée dans chacun des six plus grands SSM), et dans des directions opérationnelles pour des pôles de taille moyenne (20 à 30 experts de la donnée) à la DGFIP, DINUM, DSS, DGA, DGEFP et au pôle numérique ministériel à la transition écologique. La DGEFP fait majoritairement appel à des prestataires et n'a que très peu internalisé les compétences en dépit de besoins importants.

Au total, **une très grande part des experts de la donnée se situe en dehors des directions du numérique**, particulièrement chez les *data analysts*. Les *data scientists* se trouvent aussi majoritairement en dehors des directions du numérique.

Deux modèles d'organisation principaux coexistent : la constitution de pôles centraux de la donnée ou un modèle hybride combinant pôles et recrutement d'experts de la donnée dans les directions métiers. Le premier modèle (centralisé) se rencontre plutôt dans les ministères ou directions où il n'y a pas encore beaucoup d'experts de la donnée (hors SSM) de façon à disposer d'un écosystème *data* de taille suffisante pour se développer.

Le deuxième modèle (hybride) émerge dans des organisations qui ont atteint une certaine maturité et où une taille critique de plusieurs dizaines d'experts de la donnée est atteinte : il vise alors à placer des experts près des métiers afin de traiter de façon pertinente des cas d'usage opérationnels. En effet, l'enjeu est de conjuguer l'expertise *data* et les besoins métiers. Les pôles concentrent une très forte proportion des experts de la donnée dans les ministères du premier modèle (centralisé), plus de 80 %, et une proportion encore sensible dans les ministères du deuxième modèle (distribué) : autour de 40 % avec une tendance à la baisse dans les deux années à venir. Le cas de la DGFIP est éclairant : en plus des pôles centraux, une centaine d'experts de la donnée vont prochainement travailler localement, disséminés dans de nombreuses unités opérationnelles. En effet, sur la base de l'expérience passée, « *quand les experts data sont au sein du métier, cela permet de transformer le métier et de changer les pratiques* ».

3.2.5 Des métiers particulièrement en tension

D'après les résultats de l'enquête conduite par la mission, les métiers les plus en tension sont, en dehors de la statistique publique, ceux de *data scientist* et de *data architect* avec des écarts entre les effectifs et les besoins actuels supérieurs à 20 % des effectifs et à 20 ETP. Au sein de la statistique publique, ce sont les *data scientists* et les polyvalents.

Certains recruteurs évoquent des difficultés de recrutement liées aux salaires des ingénieurs, à l'impossibilité de trouver des *data architects* parmi les fonctionnaires ou à la pénurie de *data engineers* par rapport aux besoins du secteur privé.

Encadré 1. Les stratégies dans le secteur privé

La mission a interrogé deux sociétés de conseil, Conix et Quantmetry, afin de disposer d'un éclairage sur les organisations et les pratiques dans le secteur privé.

Si les organisations sont très variables en fonction du secteur d'activité, de la place de la donnée dans les modèles d'affaires et de la maturité des entreprises, il semble¹⁶ que la tendance dominante soit de s'orienter vers un modèle d'organisation des compétences « data » hybride avec un pôle centralisé qui a une fonction de pilotage stratégique, de coordination ou d'animation, et des compétences réparties au sein des directions métiers. Ce serait notamment le cas dans les entreprises dont le modèle d'affaires ne repose pas intrinsèquement sur la donnée. Le pôle central recouvre notamment la fonction de *chief data officer* et peut inclure une cellule d'innovation (« data lab »).

Ce modèle aurait notamment l'avantage de concilier au mieux le besoin d'identifier et de traiter les cas d'usage au plus près de l'expertise métier et celui d'une transversalité favorisant notamment la cohérence des initiatives ainsi que la mutualisation des outils et des approches.

Un point clé qui n'est pas toujours idéalement satisfait, et qui peut même être source de friction, est la bonne articulation et collaboration entre les experts *data* et la DSI, essentielles pour la réussite de l'intégration des solutions dans les applicatifs métiers.

En termes de compétences, le métier le plus en tension (pour lequel les entreprises éprouvent le plus de difficultés à recruter) n'est pas actuellement celui de *data scientist* pour lequel les formations se sont multipliées, mais celui de *data engineer*, qui constitue une compétence critique pour intégrer et industrialiser les solutions développées par les *data scientists*. On peut à cet égard penser que le besoin de compétences en ingénierie de la donnée s'accroît lorsque les entreprises gagnent en maturité et qu'elles parviennent à un stade de mise en production des cas d'usage.

3.2.6 Une forte croissance anticipée dans les deux prochaines années

À la date de rédaction de ce rapport, les ministères, pour la plupart, n'ont pas encore de feuille de route sur la *data*¹⁷. Quelques-uns disposent toutefois d'une stratégie de transformation numérique qui inclut un volet sur la donnée. Mais ils butent tous sur une difficulté : ces métiers sont très récents et peu stabilisés. Ils ne sont guère intégrés dans les outils permettant de les programmer et gérer un passage à l'échelle.

Ainsi, les effectifs actuels ne répondent pas complètement aux besoins actuels des ministères, et encore moins aux besoins dans deux ans d'après les résultats de l'enquête conduite par la mission. En effet, ils couvrent 90 % des besoins aujourd'hui.

Tableau 4 : Experts de la donnée en 2023 par métier au sein de l'État hors Insee et SSM

en équivalent temps plein	Effectifs en 2023	Evolution en deux ans	
<i>Data scientists</i>	270	+100	+59 %
<i>Data analysts</i>	400	+160	+67 %
<i>Data engineers</i>	140	+40	+40 %
<i>Data architects</i>	40	+20	+100 %
Polyvalents	80	+10	+14 %
Autres experts de la donnée	90	0	+0 %
Total experts de la donnée	1 020	330	+48 %

Source : réponses au questionnaire de la mission et estimations des auteurs

¹⁶ Voir notamment le baromètre publié par Quantmetry [baromètre des directions data 2020](#)

¹⁷ Pour mémoire, la circulaire du Premier ministre en date du 27 avril 2021 fixe au 15 juillet 2021 la date limite pour la formalisation d'une feuille de route pour chaque ministère.

Dans deux ans, le nombre d'experts de la donnée devrait augmenter fortement pour répondre aux besoins, de +20 % en globalité, et plus fortement encore en dehors de l'Insee et des SSM, de +50 %. Ils seraient alors environ 2 400 au total en 2023 sur le champ de l'enquête. Hors Insee et SSM, le métier qui devrait s'accroître le plus est celui de *data analyst* en raison d'une forte hausse prévue à la DGFIP qui compte former localement une centaine de personnes déjà en place. En dehors de la DGFIP, la hausse la plus forte sera chez les *data scientists* (+100 en deux ans). Le Ministère des Armées (DGA comprise) est probablement un autre acteur qui a besoin de rehausser significativement ses effectifs en experts de la donnée, principalement *data scientists* et *data analysts*. Au sein de l'Insee et des SSM, le nombre de *data scientists* devrait s'élever plus fortement que pour les autres métiers et d'un tiers en deux ans. Au total les administrations d'État emploieraient 430 *data scientists* sur le champ de l'enquête en 2023.

Au-delà de cet horizon, une faible proportion des ministères et des directions ont fourni un effectif cible à cinq ans : encore plus rares sont ceux qui l'ont fait en s'écartant de la cible à deux ans.

Les éléments recueillis à travers l'enquête comme les nombreux entretiens réalisés par la mission témoignent que **le processus d'intégration de la dimension data dans les organisations est encore loin d'avoir atteint sa maturité**, même si la situation est inégale selon les différents ministères. Les initiatives les plus avancées concernent notamment des services qui ont véritablement intégré l'enjeu de transformation par la donnée depuis plusieurs années au cœur de leur stratégie de modernisation, comme le Ministère des Armées ou le service du contrôle fiscal à la DGFIP. En outre, l'existence d'une culture d'ingénieur ou statistique semble avoir favorisé une certaine avance. Cela laisse présager une marge d'évolution substantielle dans les services n'ayant pas cette culture.

3.2.7 Une montée en puissance aussi chez les opérateurs

Les opérateurs ne font pas partie du champ de la mission, mais dans certains domaines la valorisation des données est largement partagée avec des opérateurs. Les opérateurs les plus concernés ont été interrogés, parmi lesquels une douzaine ont répondu au questionnaire de la mission, qui s'est entretenue avec deux d'entre eux (Health Data Hub et Pôle Emploi). Ils sont pour plus de la moitié dans le champ de la transition écologique (cf. liste en annexe 6). Parmi eux, environ 500 experts de la donnée (non compris dans les 2 000 au sein des ministères) y travaillent pour répondre à de nombreux besoins dont l'amélioration de la qualité des données, l'analyse prédictive, la cartographie et l'optimisation d'un processus.

Les experts de la donnée sont pour une bonne moitié des fonctionnaires mais la part des CDI est plus forte que dans les ministères et celles de CDD moindre en contrepartie. Les opérateurs ont recruté leurs *data scientists* parmi les fonctionnaires pour plus de la moitié et parmi des ressources externes à 30 % (prestations et chercheurs).

La répartition par métier est plus homogène chez les opérateurs qu'au sein des ministères, les *data analysts* sont certes toujours les plus nombreux et les *data architects* toujours les moins nombreux mais les écarts se réduisent : le ratio entre les deux métiers est de 1 architecte pour 3 analystes chez les opérateurs contre 1 architecte pour plus de 10 analystes dans les ministères. Les experts de la donnée sont le plus souvent regroupés dans quelques pôles – à l'instar de Pôle Emploi qui a créé il y a deux ans au sein de la direction statistique un département dédié au big data pour développer des outils d'aide à la décision à finalité opérationnelle.

Mais, la description donnée peut cacher un fonctionnement hybride. Par exemple, au Health Data Hub, une partie des *data scientists* concentrés dans l'équipe *data* sont affectés pendant plusieurs mois chez des partenaires pour traiter des cas d'usage ou pour mettre en place un processus de production de données. Ils y travaillent 80 % de leur temps et gardent un lien avec l'équipe data du Health Data Hub sur les 20 % restants.

3.3 Les limites du vivier de fonctionnaires spécialisés sur les métiers de la donnée

Les résultats de l'enquête menée par la mission font apparaître un recours très important à des profils de contractuels, particulièrement sur les métiers les plus experts. Cela peut répondre à un ensemble de raisons : l'urgence du besoin au regard du calendrier des campagnes de mobilité, le besoin de spécialistes très pointus, des incertitudes sur les besoins au-delà d'un horizon de court terme, mais aussi un vivier de fonctionnaires trop limité.

Ainsi, pour les services de l'État, hors Insee et services statistiques ministériels, ayant répondu à l'enquête, les postes de *data scientists* sont occupés à hauteur de 70 %¹⁸ par des contractuels, ce ratio s'établissant à moins de 35 % pour les *data analysts*, ces catégories représentant les profils les plus nombreux.

Parmi les corps de catégories A et A+, seuls ceux de l'Insee, corps des administrateurs et attachés statisticiens, sont spécialisés dans le domaine de la donnée. Toutefois, ces corps ne sont pas en mesure aujourd'hui au regard de leurs effectifs d'alimenter de façon significative un vivier de compétences interministériel, hors service statistique public.

Des experts sont aussi issus d'autres corps mais de façon limitée, s'agissant particulièrement du métier de *data scientist*. Pour le métier de *data analyst*, il semblerait qu'il soit plus facile de trouver des profils de fonctionnaires adéquats ou adaptables via une formation ciblée, notamment au sein des corps techniques, avec sans doute une forte variabilité en fonction du degré d'expertise attendu.

Enfin, de façon générale, il n'existe pas aujourd'hui de démarche aboutie d'identification des compétences des agents qui permettrait d'alimenter un vivier et d'organiser des parcours de formation continue pour une montée en compétences alignée sur les besoins.

3.3.1 Les corps de l'Insee

Les corps des administrateurs et attachés de l'Insee sont les seuls spécialisés dans le domaine de la donnée. Le niveau de la formation reçue et des compétences développées lors de leur parcours professionnel sont reconnues au sein de l'État.

S'ils ont une double formation en économie et statistiques, ils disposent tous, quelles que soient les options choisies lors de leur cursus à l'Ensaie et l'Ensaï, de compétences expertes dans le domaine de la donnée. Si plusieurs voies de spécialisation s'offrent à eux, particulièrement pour les administrateurs (méthodes statistiques, conseil en politique économique ou évaluation de politiques publiques), la plupart conservent après trois ou quatre postes un profil mixte de statisticien économiste qui leur permet à la fois de produire et traiter les données et d'analyser les résultats obtenus.

Le métier de statisticien évolue avec de nouvelles approches, de nouvelles méthodes et de nouveaux outils : notamment avec l'exploitation grandissante de sources administratives et le recours à des méthodes de traitement de données massives et d'intelligence artificielle. Les cursus de formation à l'Ensaie et l'Ensaï, respectivement des administrateurs et des attachés, s'adaptent en conséquence en permanence avec des cursus entre élèves fonctionnaires et non fonctionnaires qui tendent à se rapprocher.

Ainsi, si une minorité exerce aujourd'hui un métier de *data scientist* au sens strict du terme, les statisticiens des corps de l'Insee, et singulièrement les plus jeunes d'entre eux, ont la capacité d'évoluer vers ce type de métier.

Même s'ils sont concentrés sur un domaine particulier, la statistique, ces corps ont une dimension interministérielle, les administrateurs et attachés exerçant leurs compétences, non seulement à l'Insee mais aussi dans les services statistiques ministériels et dans d'autres administrations.

¹⁸ CDD, CDI, apprentis et stagiaires

Les effectifs des corps sont respectivement de 777 et 1810 pour les administrateurs (et inspecteurs généraux) et les attachés avec des flux de recrutements annuels à hauteur de 20 et 50 (y compris concours interne).

Tableau 5 : répartition des effectifs fin 2020

	Administrateurs & IG	Attachés
Insee	314	1227
GENES (écoles...)	43	134
SSM	132	259
Minefi autre	90	21
Autres ministères	16	10
Réseau régional ministère	6	76
Autre sect. Pub & div	104	43
Dispo secteur privé	72	40
Ensemble	777	1810
Recrutements annuels	20	50

Source : Insee

La part des administrateurs en fonction dans le secteur public en dehors de l'Insee et des services statistiques ministériels est non négligeable : 28 %. Toutefois, en dehors des ministères économiques et financiers pour lesquels il s'agit majoritairement d'emplois d'économistes à la direction générale du Trésor, cette part tombe à moins de 3 % pour le champ des administrations de l'État.

En ce qui concerne les attachés, seuls 8 % sont en fonction en dehors de la sphère de la statistique publique, dont une majorité dans des services déconcentrés, notamment dans les DREETS.

Les effectifs des deux corps ont connu une légère progression au cours des dernières années : + 4 % entre 2015 et 2020 pour les administrateurs, + 2,8 % pour les attachés sur la même période.

En ce qui concerne les administrateurs, cette augmentation a pour l'essentiel été absorbée par un accroissement global des effectifs affectés en SSM, qui recouvre cependant de fortes disparités entre services statistiques ministériels.

Par ailleurs, il importe de souligner que les administrateurs et attachés de l'Insee n'occupent en moyenne que 38 % des postes de catégorie A (au sens A et A+) dans les SSM. Ainsi, plusieurs services statistiques ministériels ont exprimé auprès de la mission le souhait d'accroître la part de statisticiens issus des corps de l'Insee. D'autres administrations interrogées se sont montrées intéressées par les profils Insee dont elles reconnaissent les compétences, tout en constatant qu'il s'agit de ressources rares et que se pose la question des possibilités d'évolution de carrière qu'elles pourraient offrir.

3.3.2 Les autres grands corps

Les trois grands corps d'ingénieurs qui recrutent principalement à la sortie de l'École polytechnique, en dehors du corps des administrateurs de l'Insee, le corps des Mines, le corps des Ponts des Eaux et des Forêts et le corps des ingénieurs de l'Armement ne sont pas spécifiquement positionnés sur les métiers de la donnée – ce qui n'exclut pas pour autant l'existence en leur sein de profils experts dans ce domaine.

Mais ceux-ci restent relativement peu nombreux, d'une part en raison des besoins d'ingénieurs sur d'autres spécialités et singulièrement celles qui sont attachées aux domaines particuliers d'emplois de ces corps, et d'autre part parce que d'une manière générale ils n'ont pas vocation à faire des carrières d'experts hors du cœur de métier, mais plutôt à évoluer vers des postes de *management* de haut niveau ou d'assembler.

À cet égard, lorsqu'ils détiennent des compétences dans le domaine de la donnée, celles-ci combinées à leur expertise métier les positionnent favorablement pour piloter des projets ou chantiers de transformation à forte composante *data*.

- En ce qui concerne le **corps des Mines**, sur l'ensemble des ingénieurs en activité, seuls 14 ont le terme « data » dans l'intitulé de leur poste et sont très majoritairement dans le secteur privé. Dans les nouvelles promotions, sur la vingtaine de nouveaux ingénieurs seuls un ou deux par an s'orientent vers des carrières à dominante numérique, principalement sur des problématiques de transformation par le numérique plutôt que sur de l'expertise *data*. Les premiers postes sont en outre nécessairement sur des fonctions de *manager* de terrain ;
- En ce qui concerne le **corps des Ponts, des Eaux et des Forêts (IPEF)**, sur les 21 nouveaux recrutés par an par la voie de l'École polytechnique (sur un total de 37 recrutements externes), trois à cinq suivent une formation poussée dans le domaine numérique, notamment le master « mathématiques, vision et apprentissage » (MVA) dispensé par l'ENS Paris Saclay, et poursuivent ensuite souvent par une thèse sur un domaine d'intérêt du corps et en fonction des besoins des employeurs. Une douzaine d'IPEF partent au total en thèse chaque année et une proportion très notable des thèses en cours recouvre des sujets de *data science* et d'intelligence artificielle, ou ont une composante significative dans ces domaines. Comme pour le corps des Mines, le premier poste a vocation à être un poste opérationnel en service déconcentré ;
- En ce qui concerne le **corps de l'Armement**, parmi la vingtaine d'ingénieurs recrutés à la sortie de l'École polytechnique, ceux qui s'orientent vers la *data science* et l'Intelligence artificielle font une thèse sur un sujet d'intérêt pour la Direction générale de l'armement et occupent un premier poste en lien avec leur thèse. S'il est envisagé de renforcer la composante *data science* et IA dans les spécialités du corps, cela ne devrait pas dépasser quelques nouveaux ingénieurs par an au regard des besoins à satisfaire dans d'autres domaines (le cyber de notamment). Cette inflexion ne permettra notamment pas d'accompagner la stratégie de la DGA d'augmentation rapide du nombre de ses experts *data science* et IA, pour lesquels elle continuera à avoir très majoritairement recours à des contractuels.

Tableau 6 : effectifs des corps des Mines, des IPEF et de l'armement

Corps	Effectif	Flux annuel de recrutement
Mines	1020	18
Ponts, eaux et forêts	3160	55 dont 37 externes
Armement	842	32

Source : Gestionnaires des corps

3.3.3 Les corps techniques

Deux corps d'ingénieurs ayant un positionnement interministériel ont ou développent une orientation particulière vers les métiers du numérique : le corps des ingénieurs des systèmes d'information et de communication, ISIC, géré par le Ministère de l'Intérieur, et le corps des ingénieurs de l'industrie et des mines géré par le Ministère de l'Economie, des Finances et de la Relance.

Le corps des ISIC, à vocation interministérielle, compte environ 1 000 agents avec un flux annuel de recrutement de 55 agents par concours (en 2020) et 21 au choix (en 2021), le corps étant par ailleurs alimenté par des détachements (30 en 2020).

La feuille de route DGAFP-DINUM de 2019 pour la filière NSIC a particulièrement mis l'accent sur ce corps pour développer sa contribution à l'alimentation d'un vivier spécifique pour cette filière. S'agissant des métiers de la donnée, ceux-ci n'apparaissent toutefois pas comme une spécialité significative de ce corps.

Ainsi, lors du concours de 2020 seuls trois postes étaient orientés *data*, les profils de chef de projet restant majoritaires.

Le corps des ingénieurs de l'industrie et des mines, qui compte environ 1 800 personnes en poste et recrute essentiellement des ingénieurs sur titre, a également une dimension interministérielle, ces ingénieurs occupant principalement des fonctions au Ministère de l'Economie, au Ministère de la Transition Ecologique et à l'Autorité de sûreté nucléaire. Traditionnellement positionné sur l'économie, la sécurité industrielle, les risques, le corps développe depuis peu une stratégie de réorientation vers les métiers du numérique avec une première expérience dans le cadre du concours 2020, mais sans focalisation particulière sur les métiers de la donnée. Sur une trentaine de recrutements, douze correspondent à des profils numériques dont seulement deux à trois sur des métiers de la donnée.

En dehors de ces deux corps, la mission a pu constater que dans les ministères qui disposent de corps techniques, ceux-ci permettent dans une certaine mesure de répondre aux besoins de compétences : c'est notamment le cas des ingénieurs des travaux publics de l'État (ITPE).

3.3.4 Une répartition contractuels / prestataires / fonctionnaires subie plutôt qu'organisée

Tous les avis recueillis par la mission convergent sur le fait qu'il doit y avoir un équilibre raisonné suivant une stratégie entre fonctionnaires, contractuels et recours à la prestation afin de garantir la maîtrise de cette fonction stratégique.

Dans un modèle théorique d'organisation de l'administration, le fonctionnaire est censé apporter une vision de long terme, une connaissance de l'institution et se faire le garant du travail tandis que le contractuel apporte une compétence spécialisée et que le prestataire peut gérer les aspects très pointus et dont la spécificité n'est pas propre au service public ainsi que des projets d'une durée limitée. L'équilibre permet notamment de viser une organisation efficace et d'éviter un turn-over trop important qui peut poser un risque de rupture de connaissance (voir chapitre 4).

Cependant, la situation actuelle semble relativement éloignée de ce cadre théorique.

Les fonctionnaires n'ont pas forcément la compétence pour encadrer le travail, s'étant eux-mêmes rarement retrouvés dans la situation de réaliser les tâches demandées. Les contractuels sont recrutés pour pallier un manque de main d'œuvre avec un niveau de salaire pouvant rendre difficile le recrutement de ceux ayant la plus grande valeur ajoutée. Enfin, les prestataires sont employés non pour des compétences pointues ou pour la seule période d'investissement mais de manière à pallier l'absence de ressource interne avec certains cas avoisinants la régie.

La réalité s'éloigne aussi du modèle parce que la vision des employeurs est hétérogène et que ceux-ci rencontrent des difficultés à suivre le modèle et leurs aspirations.

Les recruteurs n'ont pas un point de vue homogène : pour une partie des recruteurs, le statut du candidat n'importe guère pourvu qu'il ait les compétences, alors qu'une autre partie d'entre eux ont une préférence pour des fonctionnaires qui rassurent par leurs profils et compétences connues a priori, ou pour des CDI dans des postes demandant une continuité longue, comme des postes de chef de grands projets. Le recours aux CDD se fait de façon ambivalente : d'un côté, dans un contexte où le recrutement en CDI n'est guère l'usage ; d'un autre côté, pour combler un besoin limité dans le temps.

Face aux difficultés rencontrées par les recruteurs, la répartition entre les ressources internes et externes, et entre les statuts, est plus souvent le fruit de l'histoire et des pratiques que d'une stratégie délibérée. Quand un poste s'ouvre, le recruteur active plusieurs canaux et choisit le candidat qui correspond le plus aux compétences recherchées. L'absence de vivier de fonctionnaires suffisant oblige à chercher des solutions ailleurs.

De plus le contexte contraignant en termes d'effectifs orientés globalement à la baisse dans de nombreux endroits (plafonds d'emploi) peut inciter des administrations à une certaine prudence avant de consolider des emplois. L'appel à la prestation est considéré comme plus cher et nécessitant tout de même un investissement fort en interne mais il comporte toutefois l'avantage d'être sans impact sur le plafond d'emploi ; avantage le rendant nécessaire pour développer l'activité *data* quand le plafond est atteint.

Ainsi, dans la plupart des cas, il s'agit d'organisation opportune, une manière de faire avec les moyens du bord plutôt qu'une véritable stratégie. Même le dispositif EIG, qui porte en lui cette vision stratégique de l'emploi des contractuels venant apporter une compétence nouvelle et rare, est parfois utilisé plus comme un moyen de contourner le plafond d'emploi en dégageant des ressources supplémentaires que comme un véritable levier de transformation RH. Il est d'autres fois un sas vers l'emploi public¹⁹.

3.4 Alimenter un vivier de fonctionnaires

Pour réduire l'écart entre les besoins et les effectifs et les risques liés à un sous-dimensionnement en experts de la donnée, l'enjeu consiste à accompagner la montée en puissance des besoins en tenant compte :

- de l'absence de maturité des organisations *data* et de stratégies en cours de construction ;
- d'une identification des besoins à moyen terme qui au travers de l'enquête confirme une tendance mais reste imprécise ;
- de la diversité des besoins et des spécialités *data* ;
- du besoin de flexibilité ;
- des spécificités de l'évaluation de l'intelligence artificielle.

Face aux nouveaux besoins importants d'experts de la donnée, une première question consiste à examiner les réponses qui peuvent être apportées pour chacun des métiers à partir des corps de fonctionnaires, particulièrement pour les *data scientists* qui sont fonctionnaires pour moins d'un quart (en dehors des SSM).

3.4.1 S'appuyer sur les corps de l'Insee pour le métier de *data scientist*

Pour le métier de *data scientist*, le levier de la formation et de la qualification du personnel en place n'est envisageable que pour des personnes qui ont un très solide bagage mathématique, et le recrutement en sortie de l'École polytechnique et autres grandes écoles semble naturel.

Les corps des administrateurs et attachés de l'Insee étant spécialisés sur les métiers de la donnée, ils paraissent les plus pertinents pour développer une stratégie d'alimentation d'un vivier de *data scientists* fonctionnaires pour lesquels l'enquête montre la faiblesse des ressources actuelles. Les autres corps peuvent aujourd'hui alimenter ce vivier et continueront à le faire mais sans envisager une stratégie de spécialisation en la matière (corps des Mines, corps des Ponts, des eaux et des forêts et corps de l'Armement).

Les corps de l'Insee, de part leur vocation première, servent aujourd'hui pour l'essentiel des besoins en statistique et en analyse socio-économique, pour leur plus grande part à l'Insee et dans les services statistiques ministériels.

Aujourd'hui, les besoins de compétences dans le domaine de la donnée, qui étaient autrefois essentiellement concentrés au sein du service statistique publique (SSP), se développent considérablement en dehors comme en attestent les résultats de l'enquête conduite par la mission, même si le SSP reste très largement majoritaire. Au regard de ces besoins et des enjeux liés à une meilleure exploitation des données au sein de l'État, le rapport de la mission Bothorel a conclu dans sa recommandation n° 19 à la nécessité de

¹⁹ Après la fin de leur contrat, environ 30 % des EIG restent en tant que contractuel et 10-15 % en tant que prestataire, tous métiers et employeurs confondus.

« *diversifier les parcours des administrateurs et des attachés de l'Insee dans l'ensemble des administrations, au-delà des services statistiques ministériels, et de valoriser le travail et la carrière des agents choisissant ces parcours.* »

La proximité en termes de compétences entre *data scientists* et statisticiens permet d'envisager une évolution de ces corps vers un positionnement affirmé sur *la data science*, dans une optique d'essaimage interministériel, le corps des administrateurs ayant alors vocation à répondre aux besoins d'encadrement dans ce domaine.

Outre la satisfaction de besoins de compétences en *data scientists* une telle évolution présenterait plusieurs avantages :

- l'apport de l'expérience des membres des corps de l'Insee pour aborder les sujets éthiques et techniques auxquels sont confrontés les *data scientists* (cf. partie 2.4 et 2.5). ;
- une collaboration renforcée entre services statistiques et les autres pôles d'expertise de la donnée, favorisant la circulation des données et l'appropriation de standards communs.

Néanmoins, cet élargissement des missions des corps et cet objectif d'essaimage interministériel ne peut être atteint sans un accroissement des recrutements, certains services statistiques ministériels regrettant déjà de manquer de fonctionnaires issus des corps de l'Insee (cf § 3.3.1).

Cela passera par une hausse du nombre des élèves fonctionnaires des deux écoles formant à ces corps. D'après les calculs de la mission, l'ordre de grandeur de cette hausse pourrait s'élever jusqu'à une vingtaine par an si tous les recrutements de fonctionnaires passaient par ces écoles et si on voulait renforcer la proportion de fonctionnaires parmi les *data scientists* (cf. hypothèses et modélisation en annexe 7).

De plus, suivant leur nombre, une hausse des élèves fonctionnaires au sein des deux écoles pourrait être confrontée à des difficultés matérielles : les locaux à l'Ensaï étant proches de la saturation, il faudrait agrandir leur taille ou développer la part de l'enseignement à distance ; le nombre d'enseignants à l'Ensaë étant déjà relativement faible par rapport au nombre d'élèves, cela pourrait conduire à l'augmenter et par conséquent le nombre d'emplois. Outre ces difficultés matérielles, les écoles auront à analyser les adaptations à apporter aux cursus de formation, notamment concernant la part de la formation suivie par les élèves fonctionnaires en *data science*.

Il y a donc lieu d'étudier dans quelle proportion un élargissement est possible et dans quelles conditions ainsi que ses conséquences en termes de gestion de ces corps.

Recommandation 2 : Élargir les recrutements des corps des administrateurs et des attachés statisticiens de l'Insee pour répondre au besoin de *data scientists*

- *Étudier les conditions de cet élargissement : nombre de postes, impact pour les écoles et les cursus de formation...*
- *Valoriser la mobilité en termes de carrière et favoriser l'accès à des postes de responsabilité en dehors de la statistique publique*
- *Examiner plus généralement les conséquences de cet élargissement sur l'organisation et le positionnement des corps de l'Insee*

L'élargissement des corps à de nouvelles missions doit nécessairement s'accompagner d'une politique de gestion cohérente avec ce nouveau positionnement. Il doit s'agir :

- d'ouvrir un accès effectif à des postes de responsabilité et à des emplois fonctionnels en dehors de la statistique publique, par exemple pour encadrer dans des domaines où la transformation par la donnée est particulièrement stratégique ;

- de valoriser la mobilité en termes de carrière. Exercer des fonctions d'expert de la donnée au sein ou en dehors de la statistique publique devrait offrir a minima les mêmes perspectives en termes de promotion ;
- de faciliter des parcours mixant statistique publique et autres usages de la *data science*. L'essaiage pourrait se faire de préférence après un premier poste à l'Insee ou en SSM pour favoriser la consolidation de la formation reçue et l'apprentissage du métier ;
- de mettre à niveau ou maintenir les compétences en *data science* par la formation continue, de façon à pouvoir mobiliser les statisticiens en place au-delà des flux de nouveaux fonctionnaires.

Pour l'atteinte de ces objectifs en termes de parcours et de carrière, la gestion du corps devra s'articuler avec « le référent métier » et « l'animateur de communauté » recommandés par la mission (cf chapitre 4).

Au total, une telle évolution des corps mériterait d'être abordée sous le prisme de la conduite de la transformation au regard de l'ensemble de ses implications.

3.4.2 Passer par d'autres filières pour les métiers de *data analyst* et *data engineer*

Pour le métier de *data analyst*, le vivier de fonctionnaires peut être élargi par le biais de la formation et la qualification du personnel en place occupant des fonctions proches, et même par la reconversion de personnes plus loin du métier. La mission considère qu'il n'y a pas lieu de mobiliser les corps davantage, et recommande de poursuivre la politique de requalification et de reconversion pour répondre aux besoins en *data analysts*. Si cela ne suffisait pas, une piste serait de développer l'essaiage des attachés statisticiens vers des postes de *data analysts* hors de la statistique publique.

En revanche, pour répondre à la demande en *data engineers*, les reconversions seront plus rares et ne suffiront pas à répondre à l'évolution des besoins. Le recours aux contractuels risque d'être compliqué puisque ce métier est aussi en très forte tension dans le secteur privé. Il y a donc un intérêt à chercher à étoffer des corps en *data engineers* :

- le corps des ingénieurs de l'industrie et des mines et le corps des ingénieurs des systèmes d'information et de communication, à condition d'arriver à recruter sur titre davantage de profils adéquats. Cela pourrait passer, par exemple, par une plus grande articulation avec l'Ensaï et ses élèves non fonctionnaires en promouvant les besoins en compétence et les carrières ;
- le corps des attachés statisticiens à condition qu'il puisse recruter davantage (voir partie précédente sur la mobilisation des corps de l'Insee).

Le métier de *data architect* est encore confidentiel dans la fonction publique d'État. Quand il ne le sera plus, il sera temps de faire un bilan sur son alimentation et sa gestion. En attendant, les voies explorées pour les trois autres métiers peuvent servir ainsi que les mesures générales abordées dans le chapitre 4.

3.5 Activer l'ensemble des leviers dans une approche métier

La partie précédente s'est intéressée aux possibilités offertes par les viviers de fonctionnaires selon les métiers, mais, au regard des besoins à satisfaire, il est capital de mobiliser tous les leviers pour augmenter les chances d'y arriver : fonctionnaires experts recrutés dans les corps, contractuels, réorientation professionnelle d'agents, recours à des dispositifs tels EIG (en confortant la prime donnée aux projets comportant une intention d'internalisation) et appels à des prestataires. Ceci doit être fait en visant deux équilibres :

- Un équilibre entre fonctionnaires / contractuels / prestataires optimal et maîtrisé, avec des enjeux différents selon les métiers (*data scientists*, *data analysts*, *data engineers*). Le recours à des prestataires a vocation à se faire en complément pour des besoins ponctuels très spécialisés ou des

besoins d'une ampleur exceptionnelle, en raison des coûts et du besoin d'acculturation durable à la donnée au sein des organisations. Et il nécessite d'avoir en interne les compétences pour les piloter efficacement ;

- Un équilibre entre experts pointus, experts moins pointus et managers stratégiques. Un *data scientist* pointu va concevoir un système qui, une fois mis en place, sera géré par plusieurs experts moins pointus. Des managers stratégiques ayant des compétences dans le domaine de la donnée peuvent constituer un fort atout pour mener des projets de transformation par la donnée.

Ces nécessités et contraintes appellent des mesures RH pour faciliter le recrutement d'experts de la donnée puis leur gestion. Elles font l'objet du chapitre 4.

Chapitre 4 : Vers une gestion dynamique des compétences

Aujourd'hui, on constate une gestion des ressources humaines sans stratégie globale ni spécifique à la fois par manque d'outils et par manque de recul face à un processus d'intégration d'une expertise de la donnée qui est, hors statistique publique, largement en cours de développement.

4.1 Des compétences utilisées mais encore peu pilotées

Les métiers spécifiques et émergents comme ceux de la donnée appellent tout particulièrement une gestion par les compétences et un référentiel des métiers permettant de piloter au mieux la réponse aux besoins. Dans ce domaine, les outils manquent encore le plus souvent.

L'intégration des métiers de la donnée au sein de la famille SIC dans le Répertoire des emplois ministériels vient juste de commencer au Ministère des Armées, réputé figurer parmi les ministères les plus avancés en termes de gestion prévisionnelle des emplois et des compétences. Ces emplois sont encore absents des projections à six ans que réalise le ministère.

La difficulté à appréhender ces métiers dans leurs spécificités et variétés dans un contexte encore très mouvant et leur faible poids relatif en proportion des effectifs ministériels ne favorisent pas de fait la prise en compte de ces nouveaux métiers dans les grands systèmes de gestion RH ministériels. Néanmoins, un certain nombre d'évolutions sont à l'œuvre et des initiatives récentes ont vu le jour.

Ainsi, la Direction générale de l'administration et de la fonction publique (DGAFP) a pris en compte la nécessité de disposer d'outils pour « mieux maîtriser l'adéquation compétence requise/ compétence détenue » dans sa feuille de route pour la transformation numérique de la fonction RH « SIRH 2022 » : « *Des solutions numériques nouvelles et innovantes seront développées en appui au métier de la gestion des ressources humaines, en particulier dans le domaine de la GPEEC et de la gestion des talents, avec l'expérimentation d'outils et de suivi des compétences (via les entretiens d'évaluation ou des données déclaratives). Ces outils, reliés aux autres composants du SIRH permettront d'améliorer les parcours des agents et l'allocation des ressources humaines* ».

Le développement d'un outil de gestion des viviers, des talents et des compétences est ainsi en cours. Son objectif est de permettre aux ministères de « *collecter les compétences des agents, faire correspondre des compétences à un emploi ou à un métier et accompagner les agents dans leur parcours professionnel afin qu'ils deviennent acteurs de leur mobilité et de leur carrière professionnelle* ».

La DINUM actualise régulièrement le référentiel des métiers et compétences des systèmes d'information et de communication et par suite contribue en lien avec la DGAFP à la mise à jour du répertoire interministériel des métiers de l'État (Rime). Ces actualisations concernent à la fois un enrichissement de la nomenclature des métiers et une description affinée des compétences correspondantes. Toutefois, en ce qui concerne les métiers de la donnée, le Rime n'intègre pas encore la diversité des métiers pris en compte dans le référentiel de la DINUM : si le métier de *data scientist* y figure bien, ni celui de *data analyst* ni celui de *data engineer* ne sont répertoriés. On rappellera que le Rime a vocation à servir de référence pour la publication des fiches de postes sur le site « place de l'emploi public » et à être décliné au sein des répertoires ministériels des métiers.

Il reste qu'en tout état de cause, la diversité des spécialités et des besoins peut nécessiter une gestion des compétences à une maille beaucoup plus fine que celle du Rime ou des répertoires ministériels des métiers. Le besoin d'un expert en traitement du langage naturel nécessite ainsi par exemple d'identifier une compétence spécifique au sein de la catégorie *data scientist*.

La mission a eu connaissance de quelques initiatives qui montrent qu'une gestion plus fine des compétences commence à s'organiser à une échelle infra-ministérielle au plus près des directions employeurs :

- La Direction générale de l'armement a intégré en 2019 dans sa gestion prévisionnelle un métier « *Data science et intelligence artificielle* » pour lequel elle a notamment évalué le besoin de développement à un horizon de court et moyen terme ;
- L'État-major de l'armée de terre a également mis en place un dispositif de suivi et d'accompagnement en termes RH de la montée en puissance des besoins dans le domaine de la donnée et de l'intelligence artificielle, avec notamment une spécialisation de certains officiers et la formation de sous-officiers en vue de l'acquisition de compétences de *data analyst* ;
- Le service du numérique du ministère de la Transition écologique a engagé une cartographie des compétences de la filière numérique, en collaboration avec la DRH, sur la base de questionnaires remplis par les agents.

En matière d'adaptation des compétences et de réorientation vers des métiers de la donnée, des actions commencent ainsi à être menées. En dehors des cas cités plus haut, d'autres exemples ont pu être identifiés par la mission concernant la formation de fonctionnaires issus des corps métiers spécialisés à la DGFIP, aux Douanes ou encore à la DGAC.

Cependant, il importe de souligner que l'on n'assiste pas, à ce stade, à une généralisation de plans structurés de réorientation professionnelle, mais plutôt à des actions relevant de la gestion de cas individuels.

Ce constat tient sans doute pour partie au manque de recul et à la difficulté à anticiper les besoins futurs de façon suffisamment précise en termes de compétences attendues, comme le montre l'enquête que la mission a menée auprès des ministères. En outre, certains départements ministériels ont à ce jour des besoins relativement limités qui peuvent ne pas nécessiter l'organisation de plans de formation dédiés. **L'une des difficultés est par ailleurs l'identification précise des compétences des agents**, préalable indispensable à une réorientation professionnelle efficace, et qui requiert des outils spécifiques avec la mise au point d'un référentiel de compétences adapté. À cet égard, le service du numérique du Ministère de la Transition écologique a fait part à la mission des difficultés éprouvées pour exploiter les déclarations des agents et qualifier les compétences en l'absence d'un référentiel et d'une terminologie partagés.

Un autre frein, souligné notamment par Pôle emploi auprès de la mission, est le coût élevé de reconversion vers des métiers nécessitant une expertise spécialisée de bon niveau. Par exemple, une formation en alternance de 80 heures est mise en place pour faire monter en compétence des informaticiens vers des métiers de *data analyst*. Le coût est même considéré comme dirimant pour une réorientation vers des métiers de *data scientists* de personnels qui n'ont pas le bagage théorique nécessaire.

Pour autant, une offre de formation qui couvre un spectre large de la sensibilisation (initiation) à la professionnalisation se met progressivement en place. L'IGPDE a ainsi par exemple inscrit à son catalogue 2021 des formations courtes sur le *big data* et l'intelligence artificielle et expérimente des formations certifiantes à R et Python.

Un marché interministériel de formation aux compétences numériques porté par le Ministère des Armées avec la DGAFP, la DINUM et la DAE est en place depuis septembre 2020 et réunit huit ministères et 54 établissements publics adhérents pour des formations allant de la simple sensibilisation à l'expertise. Il comporte un lot « développement, gestion et analyse des données » avec 84 formations axées sur le *big data*, l'intelligence artificielle, la *business intelligence* et la gestion des bases de données.

Enfin, le plan France relance comporte au titre de l'axe « Innovation et transformation numérique » un volet spécifique « Professionnalisation des filières numériques publiques » qui a vocation à financer des actions visant notamment à « développer ou mettre à niveau les compétences de professionnels du numérique déjà employés par ces structures publiques » et qui cible particulièrement les métiers en tension. Quatre projets ministériels ont été retenus au premier trimestre 2021 dont un spécifiquement axé sur une montée en

compétence dans le domaine de l'expertise en données : la professionnalisation du traitement des données des économistes du Trésor via des formations avancées et certifiantes en R.

4.2 Un enjeu de fidélisation des profils d'experts de la donnée

Compte tenu de la situation encore émergente, des tailles d'équipe relativement restreintes, la majorité des responsables d'équipe de *data science* interrogés par la mission confient le plus souvent être concentrés sur une phase de constitution des équipes et n'être pas encore concernés par une phase de pérennisation. Le cas des renouvellements de CDD n'a ainsi pas été beaucoup rencontré dans le contexte *data* même si, par exemple, un service témoigne procéder à son onzième recrutement en quatre ans et pour trois postes. Il y a pourtant un vrai enjeu de fidélisation.

4.2.1 Un intérêt à prolonger les contractuels

Pour les métiers de la donnée soumis à des évolutions technologiques rapides, le fait de recourir à des contrats à durée déterminée peut relever d'une stratégie RH. En effet, l'emploi de contractuel, déjà inséré sur le marché du travail, permet d'employer des compétences « à l'état de l'art » sans avoir à attendre le temps de la formation.

Même si le prolongement de cette logique invite à renouveler l'opération et à remplacer les contractuels régulièrement, les employeurs interrogés par la mission considèrent très souvent que le départ des contractuels est plus une perte pour l'administration qu'une opportunité. Cela semble tenir à deux raisons. D'une part, au-delà des compétences techniques des métiers de la donnée, la connaissance « métier » est importante. Comprendre les données et le contexte d'un domaine de la santé ou de la sécurité par exemple nécessite une période que l'on peut qualifier de formation. Un service statistique ministériel cite par exemple des « personnes qu'on a formées et qu'ensuite on nous a piquées », formulation révélatrice de cet aspect formation.

Par ailleurs, peut-être du fait de certains écueils potentiels présentés plus haut (infrastructures informatiques) mais de façon plus vraisemblable du fait de la connaissance de l'environnement du secteur public, les experts de la donnée peuvent, comme les autres contractuels, avoir « appris l'administration ». Ce point laisse suggérer que les emplois dans la fonction publique ont des caractéristiques communes qu'on ne retrouve pas dans le privé sans que la mission n'ait pu identifier précisément cet aspect.

Enfin, comme déjà évoqué dans ce rapport, le souhait de voir les profils de qualité rester dans l'administration peut aussi être lié à la difficulté de recrutement et à l'incertitude sur la qualité des candidats. Plutôt que d'avoir à affronter l'incertitude d'un recrutement, les employeurs aimeraient pouvoir sélectionner des candidats connus et recommandés. Lorsqu'ils ont dans leur équipe un profil qui répond à leurs attentes, ils souhaitent pour lui et pour leurs collègues que ledit profil reste au sein de la fonction publique.

Dernier argument favorable au maintien des contractuels au sein de l'administration – argument qui n'est d'ailleurs pas propre aux contractuels – les personnes interrogées par la mission envisagent les mobilités au sein de la fonction publique comme un moyen de développer des synergies entre services pour le partage de données ou simplement pour l'échange de bonnes pratiques.

4.2.2 Encore trop peu de recours aux contrats à durée indéterminée

Même si la circulaire DGAFP-DINUM de 2019 l'encourage, on constate des réticences à offrir un CDI aux contractuels exerçant un métier de la donnée. Le cas échéant, le contrat à durée indéterminée est davantage présenté comme un élément de la négociation lors du recrutement plutôt que comme l'expression d'une volonté de conserver durablement la personne recrutée.

Par ailleurs, la réticence à proposer des CDI est expliquée par la difficulté à proposer des suites de carrière dans l'administration. Si le fait qu'elles devraient être ascendantes ou non ne fait pas consensus, c'est la question de la mobilité qui en règle générale est avancée comme étant le principal motif de crainte à proposer un CDI. Les mobilités au sein de l'administration n'étant pas organisées, les *managers* n'ayant pas la vision des postes à pourvoir, proposer un CDI leur paraît synonyme de figer une personne sur un poste, situation qu'ils considèrent quasi-unanimement comme sous-optimale.

Enfin, hormis depuis peu via les concours d'ingénieurs de l'industrie et de mines et d'ISIC, il n'existe pas de concours externe adapté aux experts techniques de la donnée leur permettant de rejoindre un corps de fonctionnaires.

4.2.3 Des difficultés à fidéliser les profils d'experts de la donnée

Les auditions menées notamment auprès des responsables d'équipes de *data science* au sein des ministères révèlent des difficultés importantes à fidéliser les profils d'experts, fonctionnaires et a fortiori contractuels. Force est de constater que la plupart des actions menées dans le cadre du plan d'actions sur la filière numérique DGAFP-DINUM se concentrent davantage sur l'attractivité plutôt que sur la fidélisation.

Le différentiel de rémunération par rapport au secteur privé est plus nettement marqué pour les profils les plus expérimentés. De plus, le référentiel de rémunération sur les métiers en tension du numérique a vocation à s'appliquer sur les embauches, mais ne concerne pas les renouvellements de contrats. La revalorisation des traitements et salaires est fortement encadrée (par exemple, elle est aujourd'hui limitée à 6 % au bout de trois ans pour les ministères économiques et financiers), ce qui n'est pas compétitif (et de loin) par rapport au secteur privé. Ainsi, les contractuels sont de facto incités à poursuivre leur parcours dans le secteur privé s'ils souhaitent connaître une évolution salariale rapide.

Le recours aux contrats à durée déterminée (CDD) reste encore aujourd'hui la pratique majoritaire, alors même que la circulaire DGAFP-DINUM de 2019 encourage les primo-recrutements en contrat à durée indéterminée. Dans le secteur privé et a fortiori pour des profils aussi recherchés, le contrat à durée indéterminée (CDI) est la norme. Là encore, les pratiques RH du secteur public ne sont pas de nature à attirer et fidéliser les plus expérimentés, pour lesquels le CDI est jugé incontournable, par exemple pour accéder au crédit immobilier.

La capacité à offrir des parcours et des évolutions de carrière est un autre facteur de fidélisation.

Or, la gestion des contractuels telle qu'elle est aujourd'hui pratiquée dans la fonction publique, ne permet guère d'offrir de telles perspectives à des contractuels qui souhaiteraient rester dans le secteur public après une première expérience. À la question « *que peut-on proposer aujourd'hui à un data scientist après cinq ans d'expérience dans votre administration ?* » nombre de responsables d'équipes de *data science* n'ont pas été en mesure d'apporter de réponses définitives. La voie la plus naturelle d'évolution des carrières consiste à prendre des responsabilités d'encadrement. Mais en l'état actuel, cette voie n'est pas susceptible de répondre aux attentes des contractuels en poste. Certains ne souhaitent pas accepter de responsabilités managériales par crainte de dévalorisation de leur expertise technique (a fortiori dans un domaine où les technologies et les outils évoluent très vite). D'autres contractuels, qui seraient prêts à accepter de telles responsabilités sont aujourd'hui limités par le faible nombre de postes d'encadrement disponibles.

Il s'agit d'un autre élément de comparaison avec le secteur privé : en très grande majorité, les équipes de *data science* au sein de l'administration sont des équipes de taille réduite et les postes de *lead data scientist* sont rares. Les grandes entreprises, notamment dans le secteur de la banque ou de l'assurance, ont constitué des équipes beaucoup plus conséquentes. Elles sont dès lors en mesure d'offrir de meilleures perspectives d'évolution professionnelle en interne.

La mission insiste sur **la prise en compte, dès à présent, de la fidélisation** des agents experts de la donnée. Il ressort des auditions que la connaissance du « métier » (au sens de l'activité d'un ministère) et l'apprentissage du contexte de travail spécifique au service public demandent du temps et qu'il est donc nécessaire de ne pas perdre des profils une fois cette connaissance acquise.

Le rythme des embauches de ces dernières années se traduit par une faible ancienneté dans les équipes de *data science* dans les ministères. Or, plus l'État aura pris en compte dès aujourd'hui la fidélisation de ces profils rares, plus il sera en mesure de faire face à l'évolution des besoins en compétences et expertises *data*. L'annexe 7 présente différents scénarios qui montrent que fidéliser a un impact très important sur le nombre de recrutements à opérer chaque année. Il est dommageable que des profils qui ont acquis une connaissance et une pratique du secteur public ne soient pas en mesure de poursuivre leur engagement, s'ils le souhaitent et si bien sûr cela répond aux besoins des services qui les emploient.

Plusieurs pistes ont été évoquées lors des entretiens :

- L'une d'elles consisterait à alléger les contraintes existantes en matière de revalorisation salariale, au cours et entre deux contrats. Les règles existantes, couplées au recours généralisé au CDD ont un effet délétère sur la fidélisation. De facto, les profils les plus prometteurs sont incités à partir vers le secteur privé ;
- Une autre piste serait de faciliter la mobilité, soit au sein d'un même ministère soit entre ministères. En effet, l'un des principaux facteurs d'attractivité du secteur public est la capacité à offrir des projets porteurs de sens : nombre de candidats considèrent qu'il est plus enrichissant et valorisant de contribuer à lutter contre le changement climatique ou à faciliter la prise en charge de maladies chroniques plutôt que d'optimiser la performance de messages publicitaires en ligne, par exemple. Pour fidéliser ces profils, il conviendrait donc d'être en mesure de leur proposer d'évoluer régulièrement sur de nouveaux projets en interne ou en interministériel ;
- Enfin, une dernière piste évoquée est de « garder un œil » sur les parcours des contractuels qui ont quitté le service public pour être en mesure, le cas échéant, de leur proposer de réintégrer le service public après une expérience dans le secteur privé. En l'état, cette approche apparaît encore très lointaine compte tenu des pratiques RH et de la gestion des contractuels en vigueur au sein de la fonction publique.

4.3 Recommandations

4.3.1 Traiter la spécificité des métiers de la donnée au sein de la filière numérique

Le plan d'action DGAFP-DINUM sur la filière numérique comporte plusieurs volets : attractivité, recrutement, fidélisation. Ces actions concernent aujourd'hui l'ensemble des métiers du numérique. La mission recommande de décliner ces actions de manière spécifique pour les métiers de la donnée.

Recommandation 3 : Décliner les actions du plan filière numérique sur les métiers de la donnée

Cette déclinaison doit concerner l'ensemble des métiers de la donnée identifiés par la mission, et non le seul profil de *data scientist* tel que c'est par exemple aujourd'hui le cas pour le référentiel de rémunération des métiers en tension. Dans cette optique, la stratégie d'attractivité se doit de cibler les écoles qui forment des experts de la donnée (dont l'Ensaie et l'Ensa), d'adapter le discours et de donner une visibilité accrue aux projets *data* au sein de l'État. En effet, la diversité et l'intérêt des projets *data* aujourd'hui menés au sein de l'État constituent un facteur d'attractivité fort pour les candidats. Le site de promotion des métiers du numérique pourrait utilement mettre en visibilité une sélection de quelques projets concrets et l'intérêt, pour des profils experts de la donnée, de mettre leurs compétences au service de l'intérêt général.

Les exemples ne manquent pas, qu'il s'agisse des défis EIG, des projets d'intelligence artificielle ou de *data science* menés par les administrations.

Enfin, il convient de différencier les métiers de la donnée dans le référentiel de rémunération en cours d'actualisation, en y intégrant plusieurs autres profils identifiés par la mission (dont *data analyst* et *data engineer*).

4.3.2 Développer des processus RH ciblés

L'ampleur des besoins exprimés par les administrations en matière de profils *data* impose aujourd'hui de mobiliser tous les leviers (fonctionnaires, contractuels, prestataires...) mais aussi de développer ou d'adapter les processus RH pour les métiers de la donnée. La mission considère que **les processus actuellement en place ne permettent pas de répondre efficacement** aux enjeux identifiés. Les recommandations couvrent **un large spectre des actions RH** : évaluation et gestion des compétences à une maille fine, revalorisation des rémunérations pour les profils ayant acquis une expérience au sein des administrations, renforcement de la politique de formation continue des agents dans une optique d'acquisition, de maintien des compétences voire de reconversion professionnelle.

Recommandation 4 : Mettre en place une gestion par les compétences, à une maille fine, des experts de la donnée

- *Mettre à disposition des administrations des outils d'évaluation des compétences techniques, en mobilisant le financement du plan de relance*

La mission recommande de **mettre en place une gestion par les compétences fines** pour les métiers de la donnée. Cela implique notamment d'uniformiser les référentiels de compétences et de métiers existants au sein des ministères en les détaillant à une maille plus fine, mais aussi d'être en capacité de reconnaître et d'identifier les expertises spécifiques (par exemple sur l'utilisation d'un logiciel ou sur des techniques pointues, telles que le traitement des images, ou l'analyse de réseaux). Dans cette optique, il convient de fournir aux administrations **les outils nécessaires** (tests techniques, questions, tâches de traitement et d'analyse de données) pour **qualifier, évaluer et suivre** l'évolution des compétences *data* des agents dont elles ont la charge. De tels outils pourraient utilement être mis à disposition des administrations, tant pour les recrutements (qualification des compétences techniques des candidats) que pour les évolutions des parcours et des carrières (dans une optique de mobilité et, le cas échéant, de reconversion professionnelle).

La mission recommande de mobiliser des financements du **plan de relance** pour le développement et la mise en place de ces outils.

Recommandation 5 : Permettre une meilleure revalorisation des rémunérations des experts de la donnée dans une optique de fidélisation

- *Par exemple en rendant le référentiel des rémunérations applicable au-delà des primo-recrutements*

Les règles qui encadrent aujourd'hui les revalorisations des rémunérations pendant et entre deux contrats réduisent les marges de manœuvre des employeurs publics. Sur des profils où la concurrence avec le secteur privé est forte, comme les métiers de la donnée, ces contraintes contribuent largement à la moindre fidélisation des profils, alors même qu'une énergie importante aura été dépensée pour les attirer et les former aux spécificités du secteur public. Dès lors, la mission recommande de lever les contraintes existantes. Deux approches ont été identifiées : l'une consiste à revoir les règles de gestion existantes, l'autre à adapter le référentiel de rémunération existant (pour les recrutements) pour proposer une grille d'évolution salariale qui tienne compte de l'expérience acquise par un profil.

Recommandation 6 : Renforcer la formation continue des agents des administrations aux sujets des données

- Dans une optique d'acquisition, de maintien des compétences mais aussi de reconversion professionnelle

Certaines compétences techniques liées aux données connaissent une obsolescence rapide, par exemple dans le domaine de l'apprentissage machine où **le rythme d'évolution des technologies et des environnements est rapide**. Dès lors, il convient de s'assurer que les agents en poste mettent à jour leurs connaissances et pratiques par le biais de formations continues adaptées aux profils et aux évolutions des métiers. Par ailleurs, la formation continue peut aussi s'adresser à des publics plus larges, à l'image de la formation au langage R pour les économistes de la Direction du Trésor, projet soutenu dans le cadre du plan de relance. La mission recommande donc de renforcer la formation continue des agents en mobilisant l'ensemble des leviers existants : promotion et mise en visibilité du marché interministériel existant, financements d'actions de formation à la donnée dans le cadre du plan de relance, accords avec les instituts (IGPDE...) et les écoles (GENES, Télécom Paris Tech...). Des parcours de formation en ligne, de type *Mooc*, existent aujourd'hui²⁰ mais ils pourraient être utilement adaptés pour prendre en compte les spécificités du secteur public.

Enfin, une attention particulière sera accordée à la possibilité de reconversion professionnelle pour les agents déjà en poste et qui souhaiteraient s'orienter vers les métiers de la donnée.

La fonction d'animation de la communauté des experts de la donnée et le référent métier (cf. infra) ont vocation à être parties prenantes de cette gestion des compétences au cours de la carrière.

4.3.3 Expérimenter une fonction de référent métier en interministériel

Recommandation 7 : Expérimenter une fonction de référent métier *data scientist* en interministériel

L'enquête menée par la mission révèle la **forte croissance attendue** des effectifs de *data scientists* (plus d'une centaine de postes supplémentaires dans les prochains 24 mois). Dès lors, la mission recommande d'**expérimenter une fonction de référent** sur les métiers de la donnée, en commençant par le métier de *data scientist*. Positionné à la DINUM, en relation étroite avec l'Insee et en liaison avec les autres gestionnaires des corps les plus concernés, ce référent aura pour mission de **faciliter les recrutements et les mobilités** des profils d'experts de la donnée. Pour professionnaliser ces recrutements, il aura pour mission de mettre à disposition des outils d'évaluation des compétences, de contribuer à l'élaboration de référentiels de compétences, de constituer un vivier de compétences tant au niveau des contractuels que des fonctionnaires, de promouvoir la possibilité de recruter en CDI et favoriser les mobilités entre ministères.

De par sa position interministérielle, le référent métier sera amené à **accompagner les administrations** qui le souhaitent **dans le recrutement** de profils pointus (centralisation des offres d'emploi, aide à la diffusion des offres de postes). Il pourra aussi être le point de contact des écoles, des universités et des centres de formation. Il étudiera enfin la possibilité de mettre au point un comité de domaines pour reconnaître et valoriser les expertises rares. Enfin, il assurera la liaison avec les opérateurs de l'État dans une optique de brassage des profils et des compétences entre les ministères et leurs opérateurs.

La circulaire DGAFP-DINUM de 2019 prévoit la possibilité de recruter directement en contrat à durée indéterminée. La mission recommande de recourir plus fortement à cette possibilité sur le champ de l'expérimentation car elle reste largement sous-utilisée. Il y va notamment de l'attractivité de l'employeur public face à des employeurs privés pour lesquels le CDI reste la règle.

²⁰ La plateforme FUN propose notamment des formations data conçues par l'Institut Mines Télécom ou l'Inria.

Il s'agit aussi de lutter contre l'idée, souvent répandue au sein des administrations, qu'un recrutement en CDI enferme un profil (et son employeur) sur un poste donné pour une longue période. Il convient ainsi d'inclure dans l'expérimentation la question de la mobilité des *data scientists*, notamment des contrats à durée indéterminée.

La capacité à proposer et gérer des parcours des experts de la donnée est un élément à la fois d'attractivité mais aussi de fidélisation. **La mission recommande de fluidifier les mobilités, en facilitant les parcours entre directions et ministères tant pour les fonctionnaires que pour les contractuels.**

La mission considère que cette fonction de référent métier permettrait de donner une pleine consistance à une logique de **gestion des compétences en termes de métier, en appui et au bénéfice de l'ensemble des employeurs publics**. Son intérêt ne paraît en outre a priori pas susceptible d'être remis en cause par les évolutions à venir dans le cadre de la réforme de la haute fonction publique.

Si l'expérimentation est concluante (facilitation des recrutements, promotion des mobilités interministérielles), la fonction de référent métier pourrait à terme être étendue à d'autres métiers de la donnée.

4.3.4 Développer le travail entre pairs

Hormis quelques administrations, la plupart des équipes de *data science* sont aujourd'hui de taille modeste. Il n'est pas rare de trouver des experts de la donnée isolés au sein de leur service. Cette relative dispersion des ressources au sein de l'État peut en partie s'expliquer par la maturité plus ou moins importante des administrations en matière d'exploitation des données. Cependant, cela peut constituer un facteur d'inefficience, voire potentiellement de risques, à plusieurs niveaux (cf. § 2.4 et 2.5). L'exercice du métier d'expert de la donnée nécessite de pouvoir échanger régulièrement entre pairs, qu'il s'agisse de partager des retours d'expérience ou de confronter ses méthodes et ses résultats au regard d'autres experts de la donnée. Au niveau organisationnel, la taille réduite des équipes, prises isolément, ne permet pas d'offrir aux personnels de perspective de carrière (mobilité, accès à des postes de responsabilité). Dès lors, la dimension de mise en réseau des compétences apparaît d'autant plus importante. La mission formule deux recommandations qui forment un ensemble indissociable : l'animation d'un réseau des experts de la donnée et la possibilité, pour une partie d'entre eux, de consacrer une partie de leur temps de travail à des projets communs.

D'un point de vue RH, la mise en réseau des compétences contribue aussi à répondre aux enjeux de fidélisation et de gestion des parcours au sein des administrations.

Recommandation 8 : Amplifier la mise en réseau des experts de la donnée en structurant une fonction d'animation auprès de l'administrateur général des données, des algorithmes et des codes sources et en lien étroit avec le directeur général de l'Insee

La mise en réseau des compétences recouvre de nombreuses fonctions : animer un réseau, une communauté métier, mettre en relation des acteurs aujourd'hui dispersés dans l'administration, mutualiser des outils techniques (par exemple des bibliothèques, des outils d'anonymisation ou de pseudonymisation) ou du retour d'expérience de pairs pour aborder une nouvelle technique (par exemple : le test d'un nouvel environnement de développement), faire le lien avec la recherche académique pour avoir un accès privilégié à une expertise de haut-niveau²¹, proposer une assistance aux administrations pour définir des besoins fonctionnels, formuler un cahier des charges ou piloter des prestataires.

Le Lab IA, qui assure déjà une partie de ces fonctions (réunion d'un réseau de *data scientists* autour de la présentation mensuelle de trois projets de l'administration) pourrait porter en pratique cette fonction d'animation de réseau. Il faudra organiser cette fonction en bonne articulation avec l'ensemble des pôles *data* des ministères, l'Insee, les SSM, et l'élargir à terme avec les opérateurs. Au regard du poids du service statistique public (SSP) dans l'écosystème de la donnée au sein de l'État, le directeur général de l'Insee a

²¹ L'accès à des ressources de haut-niveau pouvant, par ailleurs, constituer un facteur d'attractivité supplémentaire pour les candidats.

vocation à être étroitement associé à l'animation de ce réseau, à la fois en tant que coordinateur du SSP et gestionnaire des corps des administrateurs et attachés statisticiens de l'Insee.

La participation à une communauté interministérielle n'est pas toujours formellement encouragée par les employeurs des profils experts de la donnée. Si certains y voient l'opportunité, pour leurs *data scientists* et autres, d'enrichir leur expérience avec celles d'autres administrations, certains responsables d'équipes de *data science* ont évoqué lors des auditions le risque de perdre une partie du temps de travail d'une ressource rare et compétente. Il faut cependant souligner que l'ensemble des administrations, prises individuellement et collectivement, ont intérêt à favoriser ces coopérations et ces échanges entre experts de la donnée travaillant pour le secteur public, pour plusieurs raisons.

C'est tout d'abord l'une des pistes pour répondre au risque d'isolement des *data scientists* et à ses conséquences potentiellement néfastes sur la qualité et la fiabilité de leurs productions. Le travail interministériel offre la possibilité unique de s'entraider, de travailler ensemble et partager sur les réussites mais aussi sur les échecs et les apprentissages qui en découlent. Il contribue à l'apport d'expertise complémentaire sur les projets menés par les ministères. Enfin, l'appartenance à une communauté de pratiques est un facteur fort de fidélisation des profils.

Il nous semble dès lors pertinent d'**envisager des mécanismes plus incitatifs** en faveur de la participation à la communauté interministérielle :

Recommandation 9 : Expérimenter en 2022 la possibilité pour les experts de la donnée de consacrer au moins 10 % de leur temps de travail à des projets d'intérêt partagé à une échelle interministérielle

La mission recommande d'encourager le travail interministériel et les échanges entre pairs, en dégagant une partie du temps de travail des experts de la donnée. L'équilibre doit être recherché : si le temps de travail consacré à des projets communs est trop faible, la communauté n'aura pas les moyens de fonctionner efficacement. À l'inverse, s'il est trop important, outre un risque d'inefficience, il sera difficile de convaincre les employeurs d'y adhérer. Or, leur adhésion est indispensable à la réussite de cette expérimentation.

Pour l'année 2022, 100 experts de la donnée contribuant à des projets de *data science* (*data scientists* mais aussi autres experts y contribuant) pourraient ainsi s'entraider et contribuer à l'échange de pratiques entre pairs, à la réalisation d'outils communs ou à construire des ressources pour aider les administrations à définir, piloter et évaluer des projets de *data science*.

Il conviendra de veiller à une bonne représentation des équipes de *data science* au sein de l'État par un équilibre entre les ministères, mais aussi entre les profils issus de la statistique publique et ceux hors statistique publique (notamment au sein des directions du numérique). Cette action serait par ailleurs de nature à renforcer l'attractivité de l'État-employeur et à faciliter la mobilité entre ministères. Les administrateurs ministériels des données seront mobilisés pour identifier, au sein de leur ministère, les profils et les thèmes de travail qui méritent d'être portés en commun avec d'autres ministères et avec la DINUM.

En pratique, l'animateur du réseau aura pour fonction de mettre en œuvre ce programme interministériel, en diversifiant aussi les modalités d'animation – par exemple en proposant des formats de quelques jours autour d'une question précise ou en organisant une opération annuelle plus large, à l'image de l'évènement *Google Summer of Code*.

Un bilan de l'expérimentation pourrait être établi en fonction notamment du nombre des employeurs ayant adhéré, du nombre des experts ayant contribué, des actions menées et de la satisfaction à la fois des employeurs et des experts.

Chapitre 5 : Un rôle des services statistiques ministériels repensé

Les services statistiques ministériels (SSM)²² constituent historiquement au sein des ministères des pôles d'expertise de référence en matière de traitement et d'exploitation de la donnée. Les résultats de l'enquête menée par la mission témoignent de l'importance des compétences qu'ils réunissent dans ce domaine : avec l'Insee, ils concentrent en effet environ deux tiers des effectifs d'experts de la donnée identifiés par la mission.

La transformation à l'œuvre au sein des administrations de l'État soulève donc la question de la contribution possible des SSM à l'exploitation du potentiel des données au service de l'action publique.

Elle interroge plus généralement le rôle de ces services au sein des organisations ministérielles, et notamment du point de vue de la gouvernance des données dans un contexte où l'exploitation des sources administratives à des fins statistiques se développe fortement.

5.1 Le développement de l'exploitation des données à des fins opérationnelles interroge le rôle des services statistiques ministériels

Au-delà des missions de base de base des SSM, centrées sur la production statistique (cf § 5.1.1), de nouvelles activités émergent en réponse aux besoins des directions métiers, sans que cette aide à la décision ne s'inscrive aujourd'hui dans un cadre bien défini. Ces évolutions et la prédominance de relations distantes entre services statistiques ministériels et administrateurs ministériels des données témoignent d'un besoin de clarification du positionnement des SSM au sein des écosystèmes ministériels des données (cf § 5.1.2).

5.1.1 Rôle et compétences des services statistiques ministériels

5.1.1.1 Des missions traditionnelles exercées dans un cadre rigoureux

Au nombre de 16, les services statistiques ministériels exercent, au sein de chacun leur ministère d'appartenance (cf. annexe 8), deux missions traditionnelles, communes à tous les SSM²³ :

- d'une part, la production de statistiques publiques à partir d'enquêtes ou de sources administratives, à des fins d'information générale et pour éclairer le débat public ;
- d'autre part, l'appui aux ministères pour l'élaboration et l'évaluation des politiques publiques. Plusieurs SSM ont au surplus à ce titre explicitement une fonction de prospective.

Ces missions sont menées dans un cadre méthodologique qui offre des garanties de qualité, en particulier au regard du code des bonnes pratiques commun à tous les organismes de statistique publique de l'Union européenne²⁴. Celui-ci assure par exemple la solidité des méthodes, la qualité des données produites, leur accessibilité et leur pertinence.

²² Les services statistiques ministériels (SSM) constituent avec l'Insee le service statistique public, tel que défini par la loi 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques Services d'administration centrale des ministères, leur liste est fixée par arrêté du ministre chargé de l'économie, pris après avis de l'Autorité de la statistique publique. Cette liste figure en annexe du décret n° 2009-250 du 3 mars 2009 relatif à l'Autorité de la statistique publique

²³ [Charte des services statistiques ministériels](#), édition 2019, Insee

²⁴ [Code de bonnes pratiques de la statistique européenne](#), édition 2017

À travers ces missions et le cadre dans lequel elles sont réalisées, **les SSM acquièrent une bonne connaissance des données et développent une expertise et des compétences pointues pour leur traitement et leur interprétation** : connaissance des sources de données, de leurs limites et potentiels, capacité à juger de la représentativité des données collectées et à éviter les biais de sélection, contrôle et mise en qualité des données, maîtrise des traitements pour passer d'une donnée brute à une donnée exploitable, forte attention portée à la qualité des métadonnées, à la référence à des nomenclatures et des concepts pertinents, à la reproductibilité des processus et l'interprétation des résultats au regard de modèles.

Les études et les analyses auxquelles donnent lieu ces missions permettent en outre aux SSM de développer une bonne compréhension des phénomènes dans leur champ de compétence.

Ils utilisent les innovations, les outils et les méthodes de la science des données qui se développent, comme le traitement croissant de données non structurées, les méthodes de traitement de données massives (tickets de caisse, offres d'emploi en ligne, etc.) ou de nouvelles techniques d'appariement de bases de données (croisement de données administratives).

Les sources de données utilisées au service des missions des SSM sont nombreuses et variées :

- des données à caractère administratif collectées ou générées dans le cadre de la mise en œuvre des politiques ministérielles, par exemple des données de gestion d'une direction métier ;
- des données de même nature provenant d'autres ministères et d'opérateurs ;
- des données recueillies via une collecte statistique effectuée par le SSM lui-même, par un autre SSM ou par l'Insee ;
- des données privées si elles comportent un caractère d'intérêt général.

Les SSM ont le droit d'accéder à toutes ces données dans le cadre de leur mission à condition de respecter leur confidentialité²⁵.

5.1.1.2 Pour certains SSM, une expression plus large de leurs missions

Les textes réglementaires ministériels définissant les missions des SSM, lorsque ceux-ci ont un positionnement suffisant dans la hiérarchie organisationnelle, élargissent pour certains d'entre eux leur champ d'action à des fonctions de gouvernance et/ou d'aide à la décision.

Ainsi, par exemple :

- le SSM du Ministère des Solidarités et de la Santé, la Drees, est chargée « *d'administrer les données de santé, c'est-à-dire de développer leur usage notamment à des fins de décisions opérationnelles éclairées* » ;
- le SSM du Ministère de l'Enseignement supérieur, de la Recherche et de l'innovation, le SIES, « *conçoit, à partir de ces travaux, des outils de mesure de la performance et d'aide à la décision* » ;
- le SSM du Ministère de la Justice « *participe à la conception et apporte son expertise à la mise en place des outils de gestion et des systèmes d'information développés au sein du ministère, afin d'assurer leur adéquation à l'atteinte des finalités statistiques* » ;
- le SSM sécurité intérieure au sein du Ministère de l'Intérieur « *assure la centralisation, le contrôle, le traitement et la diffusion des données relatives à la criminalité et à la délinquance enregistrées par la police nationale et la gendarmerie nationale* » ;
- le SSM de l'Éducation nationale, la DEPP, « *coordonne la production et la mise à jour de l'ensemble des nomenclatures et des répertoires des systèmes d'information du ministère* ».

²⁵ Ils mettent à disposition des chercheurs des données individuelles dans des conditions de confidentialité maîtrisées via le Centre d'accès sécurisé des données.

5.1.2 Positionnement des services statistiques ministériels vis-à-vis des besoins d'exploitation des données au sein des ministères

5.1.2.1 De nouvelles activités émergentes

Au-delà de leurs missions traditionnelles, une bonne proportion de SSM déclarent déjà intervenir dans le cadre de **projets d'exploitation de données à visée opérationnelle** ou plus généralement **d'aide à la décision** pour des directions métiers de leur ministère. Par exemple : mise à disposition de données au niveau de chaque bâtiment, acquisition et qualification commune des données pour la satisfaction des besoins statistiques et décisionnels, chiffrage des surfaces concernées par l'interdiction d'épandage à cinq mètres des habitations, expertise de données pour la mise en place d'un dispositif d'alerte sur la déforestation importée, travaux sur les temps d'accès aux soins urgents, travaux dans le cadre de la gestion de crise liée à la Covid-19, analyse textuelle (*machine learning*) pour identifier la cyberdélinquance dans les procédures...

Ces activités peuvent correspondre à la fois à la mise en œuvre **de traitements** et la **fourniture de solutions aux directions métiers** ou à **un appui méthodologique**. Elles relèvent à la fois du champ de la statistique classique et de façon plus émergente de la *data science*. Elles s'inscrivent dans une grande gamme allant de l'acquisition de données à leur *data visualisation* en passant notamment par la participation à des projets de constitution de *hub* de données, par la mise en place de référentiels, par la qualification (fiabilisation) des données ou encore des analyses prédictives. Ces activités donnent lieu à la fourniture de données, d'algorithmes, de modélisations, d'indicateurs, de schémas de systèmes d'information...

Les finalités sont le plus souvent la fiabilisation des données, le suivi, le ciblage et le pilotage des politiques publiques. Certains SSM ont une unité spécialisée pour répondre aux demandes de données et de prestations non statistiques (aide à la décision) ; parfois cette unité leur est accolée.

En raison de leur expertise et de leur connaissance d'un grand nombre de sources de données, presque tous les SSM reçoivent en réalité des demandes de ce type, qu'ils ne sont pas toujours en mesure d'honorer.

D'une manière générale, les SSM interrogés par la mission considèrent que **le principal facteur limitant est celui des moyens en personnel**, dans un contexte où la production statistique est prioritaire, et parfois aussi un manque de capacités informatiques (matérielles ou logicielles). Dans certains cas, les compétences requises en *data science* ou en intelligence artificielle ne sont pas disponibles au sein du service statistique ou n'existent pas.

5.1.2.2 Une limite éthique et juridique à ces nouvelles activités

Les textes de niveau national et européen fondant les missions des services statistiques et le cadre dans lequel elles sont appelées à s'exercer ne font pas obstacle à ce que ceux-ci puissent réaliser d'autres activités que leurs missions de base.

Par contre, une limite juridique forte s'impose : **ces activités ne peuvent avoir pour finalité des décisions individuelles dès lors qu'elles se fondent sur des données relevant de la statistique publique.**

En effet, un certain nombre de dispositions limitent ou empêchent la réutilisation de données issues de traitements à finalité statistique ²⁶ si celle-ci doit aboutir à des décisions individuelles, autrement dit si elle répond à d'autres motifs que la production d'études ou de résultats statistiques. Ainsi, le secret statistique défini dans la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques interdit toute communication de renseignements individuels hors des services habilités. Le devoir de garantir le secret statistique est en effet la contrepartie au droit d'accès particulier à des données confidentielles dont bénéficient les SSM.

²⁶ Notamment, le règlement (CE) n° 223/2009 du 11 mars 2009 révisé du Parlement européen et du Conseil relatif à la statistique européenne, le RGPD et la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

Lorsqu'il s'agit de données n'entrant pas dans le champ de la statistique publique, ces droits et limites spécifiques n'ont pas vocation à s'appliquer, le droit pour les SSM d'accéder aux données relevant dès lors du régime de droit commun et singulièrement du RGPD.

Néanmoins, même si les SSM ont la faculté juridique d'exploiter les données à des fins de décisions individuelles, **des considérations de nature éthique peuvent faire obstacle à ce type d'exploitation** : en effet, la garantie de respect du secret statistique et sa crédibilité sont essentiels vis-à-vis des particuliers et des entreprises interrogés par enquête. Dès lors, même dans l'hypothèse où les activités statistiques et les activités ayant une finalité individuelle seraient clairement séparées, leur coexistence au sein d'un même service statistique pourrait risquer d'altérer la confiance du public, l'image d'indépendance du service statistique et de compromettre gravement la réalisation des enquêtes. Le fait qu'il puisse s'agir de décisions créatrices de droit ou d'avantages pour les personnes ciblées ne paraît pas de nature à permettre de surmonter cette difficulté.

La mission considère que **ces limites juridiques et éthiques n'empêchent nullement de développer les collaborations avec les autres services des ministères** en termes de gouvernance des données, de conseil méthodologique ou encore de travaux à visée opérationnelle sans finalité individuelle. Cependant, en ce qui concerne l'aide à la décision individuelle, **ces limites demandent à être clarifiées par une doctrine commune à l'ensemble du service statistique public**. Il doit s'agir de délimiter le type d'opérations réalisables au sein d'un large spectre qui peut aller de la collecte et la structuration des données jusqu'à l'identification de situations individuelles, en passant par la mise à disposition d'outils dont les directions métiers assurent la mise en œuvre.

5.1.2.3 Des relations le plus souvent distantes avec les administrateurs ministériels des données

Malgré leur prisme *data*, les relations entre services statistiques ministériels et administrateurs ministériels des données (AMD) ne sont guère pensées et restent, sauf exceptions, distantes.

D'une manière générale, et hors le cas particulier de la Drees, seul SSM exerçant les compétences d'administrateur ministériel des données (cf. encadré 2), ces relations sont peu formalisées et peu développées ; dans certains cas, elles n'existent même pas.

Si AMD et SSM sont souvent positionnés au sein d'un même secrétariat général ou d'une même direction d'administration centrale, les textes d'organisation ministériels n'articulent pas les rôles respectifs des AMD et des SSM, s'agissant pourtant dans les deux cas de missions transverses directement axées sur les données. La seule mention relevée par la mission concerne le Ministère de l'Éducation nationale, pour lequel la fonction d'AMD au sein de la direction du numérique pour l'éducation est présentée sur le site internet du ministère « sans préjudice des compétences de la DEPP » .

L'articulation des rôles mériterait d'être clarifiée et quelques démarches dans ce sens ont été signalées à la mission, notamment au Ministère des Solidarités et de la Santé.

Une évolution intéressante se profile aussi au Ministère de l'Intérieur, pour lequel la fonction d'AMD est organisée au sein de la direction du numérique, tandis que le service statistique de la sécurité intérieure est placé sous l'autorité conjointe des directeurs généraux de la gendarmerie et de la police nationales. Il est en effet envisagé que le service statistique ait une compétence d'administrateur des données sur les données qu'il traite et d'autres éventuellement et intègre à ce titre le réseau des administrateurs de données du ministère coordonné par la direction du numérique.

Au-delà de cet exemple, **la plupart des SSM considèrent que les collaborations avec les AMD devraient être développées** au regard du rôle qu'ils jouent en matière de collecte, de mise en qualité, de fiabilisation et d'exploitation de données et de l'expertise dont ils disposent dans ces domaines. Il y a clairement un enjeu d'efficacité collective à **exploiter les synergies potentielles** entre les compétences et les activités des SSM d'une part et les finalités poursuivies par les AMD d'autre part.

Selon les avis recueillis par la mission auprès des SSM, les sujets potentiels de collaboration avec les AMD portent sur une cartographie qualifiée des données du ministère, la structuration des données ministérielles et notamment des nomenclatures communes permettant d'assurer leur interopérabilité, l'accès du SSM aux données du ministère et du secteur privé, la valorisation innovante des contenus...

De même, il y a peu de relations ou de collaborations avec les *data scientists* hors SSM du ministère, relations qui gagneraient aussi à être développées.

Encadré 2. Positionnement des administrateurs ministériels des données

L'enquête menée dans le cadre de la mission met en lumière la grande diversité des profils, des missions et des positionnements des administrateurs ministériels des données au sein des administrations.

Le plus souvent, dans les ministères, la fonction d'Administrateur ministériel des données (AMD) est organisée au sein de la direction, la délégation ou le service chargé du numérique. Trois exceptions sont néanmoins à relever.

Celles-ci concernent en premier lieu le Ministère des Solidarités et de la Santé avec la situation singulière de la Drees, seul SSM exerçant la responsabilité d'administration ministérielle des données. Ainsi, la direction du numérique, commune aux ministères chargés des affaires sociales et du travail, porte-t-elle la fonction d'AMD exclusivement sur le champ travail.

Au Ministère de la Transition écologique, la fonction d'AMD est exercée en pratique au sein du Commissariat général au développement durable (CGDD) par le laboratoire d'innovation Ecolab du service de la recherche et de l'innovation. Il est à noter que cette fonction était auparavant dévolue au service statistique ministériel, le SIES, le transfert ayant eu lieu au début de 2020.

Au Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, l'administrateur général des données est placé auprès du directeur général pour la recherche et l'innovation et de la directrice générale de l'enseignement supérieur et de l'insertion professionnelle.

5.2 Les évolutions à l'œuvre dans les instituts de statistique étrangers

La multiplication des bases de données dans l'administration, le besoin grandissant d'exploiter ces données au service de l'action publique et d'en permettre une large réutilisation interrogent le positionnement des instituts de statistiques à l'étranger comme en France.

Des réflexions sont en particulier en cours au niveau européen pour ce qui concerne les instituts nationaux de statistiques et leur positionnement au sein de l'écosystème public de la donnée, avec en particulier la question d'une évolution possible vers une fonction de *data steward* dans une acception plus ou moins large en termes de management des données et de compétences.

Telle que la définit notamment David Plotkin dans son ouvrage « *Data stewardship, an actionable guide to effective data management and data governance* », la fonction de *data stewardship* renvoie notamment à la gouvernance opérationnelle des données avec une attention particulière portée à la pertinence des métadonnées et la qualité des données.

L'une des questions majeures est celle du rôle que les instituts de statistique (INS) peuvent jouer vis-à-vis des autres administrations pour favoriser une meilleure exploitation de la valeur des données qu'elles détiennent. L'un des points critiques est de distinguer « *ce qu'ils peuvent faire* » de « *ce qu'ils peuvent aider à faire* » dans le cadre de leurs missions et pour ne pas risquer d'altérer la confiance du public dans les statistiques produites ; cette question renvoyant à celle des limites éthiques évoquées au paragraphe 5.1.2.2.

Les instituts européens ont notamment présenté leurs réflexions et stratégies lors d'un séminaire organisé sous l'égide d'Eurostat à la fin du moins de mars 2021.

Une large convergence de vues s'est exprimée en faveur de la **promotion vis-à-vis des autres administrations de standards et de concepts** permettant d'assurer la qualité et l'interopérabilité des données (notamment référentiels, métadonnées) et par là même d'une **contribution clé des INS à la gouvernance des données** à l'échelle de l'administration. Il s'agit à la fois d'exploiter les compétences des services statistiques au bénéfice de l'ensemble des acteurs et d'offrir à ces services de meilleures conditions d'exploitation des données administratives à des fins de statistique publique, dans un contexte de développement de l'utilisation de ces données par les INS.

Mais des différences d'approche sont également apparues sur les limites ou au contraire l'étendue du rôle que les INS pourraient avoir vocation à jouer :

- Certains INS vont ainsi plus loin avec un rôle de coordination et de gouvernance d'ensemble des données de l'administration à l'instar de l'Estonie. En Lituanie, dans le contexte de crise du coronavirus, la législation de ce pays a été modifiée pour permettre à l'INS de produire l'ensemble des statistiques pertinentes via la gouvernance du système d'information des données de l'État ;
- À contrario, d'autres INS soulignent des risques de perte d'identité et de confiance si l'INS gère à la fois les infrastructures de données aux fins de statistique publique et aux fins de décisions publiques ;
- Enfin, d'autres, comme en Irlande ou en Italie, considèrent que la fonction de « data steward » peut inclure la mise à disposition de compétences auprès des administrations, notamment dans le domaine de la science des données et par détachement de statisticiens.

En dehors de l'Europe, l'exemple australien fournit une autre illustration du rôle que les instituts de statistiques peuvent être amenés à jouer pour exploiter les données administratives au service du pilotage des politiques publiques.

L'Australie a lancé, dès 2015, le projet MADIP (*multi-agency data integration project*) porté par le bureau australien de statistiques (ASB) et associant huit départements ou agences gouvernementaux. Dans le cadre de ce partenariat, ABS est chargé de construire un système d'information combinant les informations publiques sur la santé, l'éducation, les paiements gouvernementaux, l'impôt sur le revenu, l'emploi et la démographie (y compris le recensement) pour créer une image complète de l'Australie dans le temps. L'objectif est de contribuer à l'élaboration, au ciblage et à l'évaluation des politiques publiques par l'exploitation de ces données de référence, combinées le cas échéant à d'autres données en fonction des sujets d'étude.

5.3 Recommandations

Au regard des enjeux de la politique de la donnée au sein de l'État, **il convient d'exploiter au mieux les synergies possibles** entre la satisfaction des besoins des directions métiers et la production de statistiques publiques et de **tirer parti des compétences et de l'expérience** des services statistiques ministériels, quand bien même leurs missions premières poursuivent d'autres finalités que l'aide à la décision opérationnelle.

Ces synergies reposent notamment sur l'utilisation de sources administratives de données et de processus de collecte de nouvelles données pouvant à la fois servir des finalités statistiques et des besoins d'outils d'aide à la décision.

5.3.1 Promouvoir une offre de services

D'ores et déjà, comme souligné au paragraphe 5.1.2.1, certains SSM répondent à des demandes de traitement de données à des fins d'aide à la décision ou d'appui méthodologique, sans que cette activité soit nécessairement reconnue comme partie intégrante de leurs missions. **Il convient de mieux structurer, développer et faire connaître ce rôle d'appui, dans un positionnement repensé de ces services.**

Les feuilles de route ministérielles dont la circulaire du Premier ministre en date du 27 avril 2021 relative à la politique de la donnée, des algorithmes et des codes sources, appelle à la mise en place rapide, doivent donc pleinement intégrer cet enjeu.

Recommandation 10 : Développer une offre de service des services statistiques ministériels adaptée au contexte particulier de chaque ministère

- *Prendre en compte cette offre de service dans la mise en œuvre des feuilles de route ministérielles autour de la donnée, des algorithmes et des codes sources*
- *Adapter les textes d'organisation pour asseoir un positionnement plus large de ces services en termes d'exploitation de données*
- *Élaborer une doctrine commune à tous les SSM sur leur champ possible d'intervention au-delà de leurs missions traditionnelles*
- *Desserrer la contrainte de moyens en lien avec cette offre de service*

Il ne s'agit pas bien entendu de développer cette offre de services au détriment des missions traditionnelles des SSM, mais bien d'élargir leur champ d'action et d'**en tirer les conséquences en termes de moyens**, en tenant pleinement compte **des gains d'efficacité** qu'il peut y avoir à exploiter des **compétences internes expertes**, familières des données et des problématiques ministérielles, qui ont un recul et une expérience quant à la compréhension des phénomènes et maîtrisent les méthodes et techniques, plutôt que de faire appel à des compétences extérieures (contractuels ou prestataires).

Cette offre de service doit permettre de clarifier auprès des autres directions des ministères, les rôles que peut et ne peut pas jouer le SSM mais aussi de présenter les SSM comme un soutien aux autres directions et non comme des entités à part. De plus elle aura été définie en impliquant la direction dont fait partie le SSM.

Comme l'a montré notamment le chapitre 3 de ce rapport, et au regard des domaines de compétences des différents SSM, les différences d'un ministère à l'autre ne peuvent s'accorder avec un modèle unique. Dès lors, en fonction des besoins et des situations, cette offre pourra recouvrir dans des proportions variables appui méthodologique, mise en place d'outils et traitements effectifs dans le cadre de projets. Elle devra en tout état de cause s'inscrire dans une étroite collaboration avec les directions d'administration centrale.

Enfin, une adaptation des textes d'organisation ministériels définissant les missions des SSM, au sein de la majorité desquels le terme « données » est absent, serait l'occasion de l'introduire et de mettre pleinement en lumière la compétence et les attributions des SSM dans ce domaine.

5.3.2 Clarifier les rôles et développer les collaborations avec les AMD et les services chargés du numérique

La circulaire du Premier ministre en date du 27 avr 2021 relative à la politique de la donnée, des algorithmes et des codes sources, appelle l'ensemble des ministères à « *mettre en place l'organisation la plus adaptée pour favoriser les synergies entre les services chargés du numérique dans (...) chaque ministère, notamment les services statistiques* ».

À cette fin, le Premier ministre demande la nomination d'un administrateur ministériel des données dans chaque ministère au 15 mai 2021. Leur rôle premier est ainsi défini par la circulaire : les AMD sont chargés d'élaborer la stratégie ministérielle de la donnée, de coordonner les parties prenantes et d'être le point de contact des utilisateurs des données et des applications numériques.

La circulaire met l'accent sur la circulation et le partage de la donnée, dans la continuité des recommandations du rapport de la mission Bothorel²⁷. Elle ne définit cependant pas de manière précise le rôle que les AMD sont amenés à jouer en matière de **gouvernance de la donnée**, dont la circulation et le partage ne sont que l'une des facettes. En particulier, il conviendra de préciser, au sein de chaque ministère, la répartition des rôles et des responsabilités en privilégiant une étroite collaboration entre AMD, services du numérique qui notamment gèrent les infrastructures de données et dans la majorité des cas portent la responsabilité d'administration générale des données, et SSM.

Recommandation 11 : Formaliser un cadre de collaboration renforcé des services statistiques ministériels avec les AMD et les services chargés du numérique

Cette collaboration a vocation à porter à la fois sur la gouvernance des données et sa mise en œuvre opérationnelle, notamment cartographie des données, constitution de métadonnées, nomenclatures, mise en qualité des données, mais aussi l'ouverture des statistiques produites par les SSM et les infrastructures de données.

Elle doit relever d'une double optique : à la fois faire bénéficier l'ensemble des utilisateurs des données des compétences développées par les SSM et garantir la bonne prise en compte des besoins des SSM. S'agissant de ce dernier point, il importe que les systèmes d'information et la structuration des données administratives soient autant que possible compatibles avec les finalités statistiques et qu'à cet effet les SSM soient informés des projets et puissent appréhender leurs impacts potentiels pour leurs activités. En outre, une bonne collaboration entre SSM et services chargés du numérique est nécessaire pour la mise en production des modèles et solutions développés par les SSM. **À ce titre, une association étroite des SSM aux projets de refonte de systèmes d'information** et notamment de mise en place de lacs de données devrait être systématisée.

En ce qui concerne l'ouverture des données, les données statistiques s'inscrivant dans un cadre spécifique avec notamment des exigences particulières en termes de qualité et de respect du secret statistique, il convient que les services statistiques ministériels conservent dans un dialogue ouvert avec les AMD et dans une contribution active à la politique d'ouverture et de circulation des données publiques, la **maîtrise de la diffusion des données statistiques**.

5.3.3 Ouvrir des possibilités de mobilité des experts des SSM vers les pôles ministériels ou interministériel de *data science*

Les services statistiques ministériels peuvent répondre aux besoins des directions métiers et favoriser une mise en œuvre efficiente des stratégies *data* non seulement en apportant un appui méthodologique et opérationnel mais aussi **en contribuant à la diffusion de compétences expertes au sein de l'État**.

C'est le sens de la recommandation 19 du rapport « *Pour une politique publique de la donnée* » de la mission Bothorel « *Diversifier les parcours des administrateurs et des attachés de l'Insee dans l'ensemble des administrations, au-delà des services statistiques ministériels, et valoriser le travail et la carrière des agents choisissant ces parcours* ».

²⁷ Le rapport Bothorel appelle à un élargissement et un renforcement du rôle des AMD. Ces derniers doivent non seulement porter la politique de la donnée au sein de leur ministère, mais aussi le sujet des algorithmes et des codes sources. Afin d'asseoir leur légitimité, le rapport appelle à doter les futurs AMDAC d'une lettre de mission signée par les ministres concernés, après consultation des directions générales et de la DINUM et en s'assurant qu'ils disposent des moyens suffisants pour l'exercice de leur mission.

En lien et dans le prolongement des recommandations formulées par la mission aux chapitres 3 et 4 concernant la mobilisation des corps de l'Insee et une plus grande fluidité et diversité des parcours et des mobilités, et au regard des besoins de compétences de *data scientists*, il conviendrait de favoriser les possibilités d'essaimage vers des pôles de *data science* et en particulier vers des postes à responsabilité et d'encadrement de profils plus juniors, notamment de jeunes contractuels non encore acculturés à l'environnement et aux problématiques publiques.

Recommandation 12 : Ouvrir davantage les possibilités de mobilité des experts des services statistiques ministériels en dehors du service statistique public, en privilégiant les pôles de *data science* et l'accès à des postes de responsabilité

Il s'agit pour ce faire de mettre en place une collaboration étroite entre les DRH ministérielles, les AMD, les SSM et l'Insee en tant que gestionnaire des corps des administrateurs et des attachés statisticiens, prenant notamment appui sur une gestion prévisionnelle des emplois et des compétences. Le référent métier proposé par la mission (cf. recommandation 7) aurait dans cette perspective un rôle central.

Ces mobilités qui concourront à enrichir les parcours et à renforcer l'attractivité des métiers de la donnée au sein de l'État, présenteront aussi l'avantage de permettre de diffuser plus largement au sein de l'État une culture de la qualité des données dans ses différentes dimensions qui fait parfois défaut et qui est pourtant essentielle à une exploitation efficiente des données. Elles **seront aussi de nature à favoriser les collaborations** entre SSM, directions métiers et services chargés du numérique dans une compréhension commune des enjeux.

Liste des annexes

Annexe 1 : Lettre de mission.....	62
Annexe 2 : Liste des personnes rencontrées.....	65
Annexe 3 : Bibliographie.....	70
Annexe 4 : Méthodologie du traitement des offres de la place de l'Emploi Public.....	71
Annexe 5 : Questionnaires de la mission.....	72
Annexe 6 : Liste des répondants à l'enquête.....	75
Annexe 7 : Modélisation du nombre de recrutements dans les années à venir.....	76
Annexe 8 : Organisation des services statistiques ministériels.....	78
Annexe 9 : Les compétences « données » des directions du numérique dans les ministères	79
Annexe 10 : Glossaire.....	81

Annexe 1 : Lettre de mission



**PREMIER
MINISTRE**

*Liberté
Égalité
Fraternité*

Paris, le 10/02/2021

Lettre de mission

Evaluation des besoins de l'Etat en compétences et expertises en matière de donnée

Le directeur de la DINUM

Le directeur général de l'INSEE

A

M. Simon Chignard, chargé de mission auprès de la chef de département Etalab

Et

Mme Danielle Bourlange, inspectrice générale de l'INSEE

M. François Brunet, inspecteur général de l'INSEE

Big data, intelligence artificielle, transformation des politiques publiques, aide à la prise de décision stratégique et opérationnelle, open data, nouvelles études sectorielles : la donnée est au cœur de la modernisation de l'Etat.

Cette montée en puissance s'est traduite 1) dans la constitution du programme TECH.GOUV d'accélération de la transformation numérique de l'Etat, 2) dans le rapport Bothorel, qui incite à se saisir des opportunités liées à l'ouverture et de l'exploitation des données pour améliorer le service proposé aux citoyens, et 3) dans la plupart des chantiers de réorganisation des administrations, qui renforcent la place de la donnée dans l'organisation et la gouvernance interne entre les services concernés (directions « métiers », services statistiques ministériels, directions du numérique). Elle se retrouve également dans les réflexions menées au sein de la statistique publique sur le positionnement et les carrières des data scientists.

Néanmoins, ces enjeux ne sauraient être appréhendés sans une évaluation qualitative et quantitative des expertises associées, et une réponse à court et moyen terme permettant d'y répondre. Faute de cela, l'Etat prendra le risque d'externaliser excessivement ces fonctions essentielles, et réduire sa capacité d'innovation. Cette analyse nécessite de bien cerner les besoins à satisfaire et les principales finalités.

La mission qui vous est confiée consiste à :

- Réaliser une typologie des besoins fonctionnels et des besoins de compétences dans chacun des ministères, à la fois quantitativement (dans les grandes lignes) et qualitativement : en d'autres termes, préciser la demande
- Clarifier le positionnement de l'Insee et des services statistiques ministériels vis-à-vis de ces besoins d'expertise, qui peuvent relever de la statistique publique ou de l'aide à la décision
- Faire le point sur les dispositifs existants permettant de recruter, au sein de l'Etat, des experts de la donnée et de la statistique (corps des administrateurs et attachés de l'INSEE, contractuels, dispositifs innovants tels les Entrepreneurs d'Intérêt Général), et déterminer le vivier disponible dans la fonction publique : quantifier l'offre
- Evaluer la capacité des corps de l'Etat (INSEE, Mines et Armement principalement) à renforcer ce vivier par primo-accession
- Enfin, analyser la capacité des instituts de formation (ENSAE, ENSAI, formation continue, IGPDE, etc.) à proposer des dispositifs de réorientation professionnelle permettant de renforcer cette capacité propre de l'Etat.

Vous veillerez à prendre l'attache des parties prenantes interministérielles et ministérielles – directeurs du numérique et réseaux des administrateurs ministériels des données et services statistiques ministériels – afin de collecter leur vision, identifier la stratégie qu'ils mettent en place pour la mettre en œuvre, et expliciter leurs attentes éventuelles en matière de filière RH.

Le rapport devra nous être remis mi-mai.

Le directeur interministériel du numérique
Nadi BOU HANNA



Le Directeur Général de l'INSEE
Jean-Luc TAVERNIER





NOTE DE SERVICE

A l'attention de

Monsieur le Directeur interministériel du numérique

Dossier Suivi par :
RIVIERE Pascal
Tél : 01.87.59.50.66
Mél : pascal.riviere@insee.fr

Montrouge, le 12 février 2021
N° 2021_12/DG75-B001
Classement n° 2021_2

Objet : Mission relative à l'évaluation des besoins de l'Etat en compétences et expertises en matière de donnée

Par courrier en date du 10 février 2021, les Services du Premier ministre ont demandé à la direction interministérielle du numérique et à la direction générale de l'Insee de diligenter une mission relative à l'évaluation des besoins de l'Etat en compétences et expertises en matière de donnée.

Monsieur Alexis EIDELMAN, administrateur hors classe à la direction de l'animation de la recherche, des études et des statistiques, est désigné comme intégrant l'équipe de mission.

Le chef de l'Inspection générale

Signé : Pascal RIVIÈRE

Pour information :

- Madame la directrice de l'animation de la recherche, des études et des statistiques
- Monsieur Alexis Eidelman, direction de l'animation de la recherche, des études et des statistiques
- Monsieur le directeur général de l'Insee

Annexe 2 : Liste des personnes rencontrées

Secrétariat Général du Gouvernement

Direction interministérielle du numérique (DINUM)

M. Nadi BOU HANNA, Directeur

Mme Charlotte CADOR, Mission Talents

Mme Soraya SAA, Mission Talents

Mme Soizic PENICAUD, Etalab, programme EIG

Mme Raphaëlle ROFFO, Etalab, programme EIG

Ministère de la Transformation et de l'Action publique

Direction générale de la Fonction publique (DGAFP)

M. Nicolas de SAUSSURE, Chef du service du pilotage des politiques de ressources humaines

M. Adrien FRIEZ, Sous-directeur des études, des statistiques et des systèmes d'information

Mme Cécile LOMBARD, Sous-directrice des compétences et des parcours professionnels

M. Emmanuel BROSSIER, Chef du bureau de la gestion prévisionnelle, des filières métiers et de l'accompagnement personnalisé

Ministère de l'Agriculture et de l'Alimentation

Conseil général de l'alimentation, de l'agriculture et des espaces ruraux

Mme Edith VIDAL, Secrétaire générale

M. Xavier ORY, Conseiller auprès du Vice-président

Service des ressources humaines

M. Grégoire JOURDAN, Expert de haut niveau en charge du pilotage du recrutement, de la formation et des parcours des ingénieurs des ponts, des eaux et des forêts

Service de la statistique et de la prospective

Mme Corinne PROST, Cheffe du service

Mme Juliette FOURCOT, Cheffe du bureau des méthodes et de l'information statistiques

Ministère des Armées

État major de l'Armée de Terre

Général Claude CHARY, Officier général du numérique et de l'innovation, Pôle numérique et innovation,

Colonel Roger BARRAU, Chef d'état-major de la direction des ressources humaines de l'Armée de Terre

Secrétariat général pour l'administration

Commissaire général Frédéric STERNENBERG, Sous-directeur du pilotage des ressources humaines militaires et civiles

Mme Maroussia RENUCCI, Adjointe au sous-directeur, sous-direction du pilotage des ressources humaines militaires et civiles

Mme Carole EL BAGGARI, Chef du bureau de la gouvernance des données RH

Mme Valérie NAVELLOU, Directrice de projet « Formation » auprès du directeur des ressources humaines

Mme Suzanne PARROT-SCHADECK, Adjointe au responsable de la fonction transformation de la fonction ressources humaines

Capitaine de corvette Rémy BALME, direction des ressources humaines

M. Olivier LANDOUR, Chef de la mission d'aide au pilotage

Conseil général de l'Armement

M. Hervé GUILLOU, Vice-président

Mme Mireille CARLIER, Secrétaire générale

Direction générale de l'Armement

M. Jérôme LEMAIRE, Chargé de mission auprès du délégué général pour l'armement

M. Mike BARGAIN, Responsable du pôle numérique à la direction technique

Autres services

M. Vincent BIAUSQUE, *data scientist*

Ministère de l'Economie, des Finances et de la Relance

Institut national de la statistique et des études économiques (Insee)

M. Jean-Luc TAVERNIER, Directeur général

Mme Karine BERGER, Secrétaire générale

Mme Sylvie LAGARDE, Directrice de la méthodologie et de la coordination statistique et internationale

Mme Françoise MAUREL, Inspection générale de l'Insee

Mme Elise COUDIN, Responsable du SSP Lab

Inspection générale des finances

Mme Florence GOMEZ, Inspectrice des finances

Mme Alice SCHOENAUER SEBAG, Responsable du pôle *data science*

Direction générale des douanes et droits indirects

Mme Isabelle BRAUN-LEMAIRE, Directrice générale

M. Amaury DECLUDT, Délégué à la stratégie

M. Christophe de SAINT-VIANCE, Administrateur des données

Direction générale des finances publiques

M. Gilles TAUZIN, Délégué à la transformation numérique

M. Lionel PLOQUIN, Administrateur des données

M. Régis TISSIER, Adjoint au responsable du pôle données

M. Stéphane CREANGE, Sous-directeur du contrôle fiscal, pilotage et expertise juridique

M. Philippe SCHALL, Chef du bureau programmation des contrôles et analyses des données, sous-direction du contrôle fiscal

M. Denis BOISNAULT, Chef du département des études et statistiques fiscales

M. Christophe BELLEGO, Chef de pôle au département des études et statistiques fiscales

Tracfin

Mme Delphine LE, Responsable du pôle *data science*

Direction générale des entreprises

M. Patrice GRELICHE, Responsable du pôle pilotage des corps techniques

M. Jean RIBEIL, Directeur régional de l'Economie de l'Emploi du Travail et des Solidarités de Bourgogne-Franche Comté

Mme Catherine MOREAU, Cheffe du bureau du pilotage des corps techniques

Mme Marie-Gaëlle PINART, Adjointe à la cheffe de bureau du pilotage des corps techniques, responsable du suivi des parcours professionnels des corps techniques

Pôle d'expertise de la régulation numérique (Peren)

M. Nicolas DEFFIEUX, Directeur

Conseil général de l'économie, de l'industrie, de l'énergie et des technologies

M. Luc ROUSSEAU, Vice-président du Conseil général de l'Economie

M. Godefroy BEAUVALLET, Chef du service de gestion du corps des Mines

Secrétariat général

M. Stéphane TRAINEL, Administrateur ministériel des données des ministères économiques et financiers

Institut de la gestion publique et du développement (IGPDE)

Mme Virginie MADELIN, Directrice

Mme Nathalie VAGLIO, Responsable du département Innovations technologiques et e-formation

Ministère de l'Enseignement supérieur, de la Recherche et de l'innovation

Mme Isabelle KABLA-LANGLOIS, Sous-directrice des systèmes d'information et des études statistiques

Ministère de l'Intérieur

Direction des ressources humaines

M. Thomas FAUCONNIER, Sous-directeur du recrutement et de la formation

Mme Claire WANDEROILD, Adjointe au sous-directeur du recrutement et de la formation, chef du bureau du pilotage de la politique de formation

Direction de la Gendarmerie nationale

Mme Marjory CANONNE, Responsable du DatatLab au STSI2

Ministère de la Justice

Sous-direction de la statistique et des études

Mme Christine CHAMBAZ, Sous-directrice

Service du numérique

M. Fabien ANTOINE, Directeur de projet « stratégie data »

Ministère des Solidarités et de la Santé, ministère du Travail, de l'Emploi et de l'Insertion

Inspection générale des affaires sociales

Mme Delphine CHAUMEL, Inspectrice des affaires sociales

Direction du numérique

Mme Hélène BRISSET, Directrice, administratrice ministérielle des données du Ministère du Travail, de l'Emploi et de l'Insertion

M. Vincent GAILLOT, Directeur de cabinet de la directrice

Direction de la recherche, des études, de l'évaluation et des statistiques (Drees)

M. Fabrice LENGART, Directeur, administrateur ministériel des données du Ministère des Solidarités et de la Santé

Mme Lucile OLIER, Cheffe de service, adjointe au directeur

Mme Charlotte GEAY, Cheffe du laboratoire innovation et évaluation en santé

Ministère de la Transition écologique

Service du numérique

Mme Anne JEANJEAN, Cheffe du service

Commissariat général au développement durable

M. Thierry COURTINE, Chef du service de la recherche et de l'innovation

Mme Béatrice SEDILLOT, Cheffe du service des données et études statistiques,

M. Thomas COTTINET, Directeur du projet Ecolab, Commissariat général au développement durable

Conseil Général de l'Environnement et du Développement Durable

M. Denis PRIOU, Membre permanent du Conseil

Opérateurs de l'État

Health Data Hub

Mme Stéphanie COMBES, Directrice

Pôle emploi

Mme Hélène CAILLOL, Adjointe au directeur, direction des statistiques, des études et de l'évaluation

M. Samuel LACHAUD, Responsable du pôle développement des compétences à la DSI

Groupe des écoles nationales d'économie et de statistique (GENES)

Mme Catherine GAUDY, Directrice

M. Laurent DAVEZIES, Directeur des études, Ensaé

Ronan LE SAOUT, Directeur adjoint, directeur des études, Ensai

Télécom Paris

Mme Ons JELASSI, Chargée d'enseignement, directrice de Télécom Evolution

Sociétés de conseil

CONIX

M. Joël BIZINGRE, Directeur du développement

Quantmetry

M. Karl NEUBERGER, Associé

Annexe 3 : Bibliographie

- Administrateur général des données : [rapport au Premier ministre sur la gouvernance de la donnée dans les administrations](#), La documentation française, avril 2018
- Apec, [Les métiers de la donnée](#), octobre 2020
- Berthold, Michael, *What does it take to be a successful data scientist ?*, Harvard Data Science Review, 2019
- Cigref, [Nomenclature des métiers du système d'information](#), juillet 2018
- Conseil économique, social et environnemental, *Économie et gouvernance de la donnée*, février 2021
- Corbett-Davies, Sam et Goel, Sharad. *The Measure and Mismeasure of Fairness : A Critical Review of Fair Machine Learning*. 2018, arXiv:1808.00023
- Dares, [Data scientists, community managers... et informaticiens : quels sont les métiers du numérique ?](#), Aurore Desjonquères, Claire de Maricourt, Christophe Michel, *L'économie et la société à l'ère du numérique*, édition 2019
- Direction générale de l'administration et de la fonction publique (DGAFP), [Répertoire interministériel des métiers de l'État](#), 2018
- Direction générale de l'administration et de la fonction publique, « *Transformation de la fonction recrutement, éléments préparatoires au plan interministériel de transformation de la fonction recrutement* », édition 2020
- France Stratégie, [Vision prospective partagée des emplois et des compétences – la filière numérique](#), Réseau Emplois Compétences, en partenariat avec le CEREQ, 2017
- Ministère des Armées, [Stratégie de l'intelligence artificielle au service de la défense](#), 2019
- Ministère de l'Économie, des Finances et de la Relance, [Faire de la donnée un actif stratégique du MEFR](#), feuille de route 2020-2021
- Ministère de la Transition Ecologique, [La donnée : un nouveau bien essentiel pour la transition écologique et solidaire et la cohésion des territoires](#), feuille de route 2018-2019
- Observatoire des métiers du numérique, de l'ingénierie, du conseil et de l'évènement (OPIIEC), [Cartographie des métiers](#)
- O'Neil, Cathy, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy.*, New York, Crown, 2016
- Plotkin, david, « *Data stewardship, an actionable guide to effective data management and data governance* », Academic Press, 2013
- Rapport [Pour une politique publique de la donnée](#), mission Bothorel, décembre 2020
- Rapport [Stratégie de ressources humaines de la filière numérique et Système d'Information et de Communication \(SIC\) de l'État](#), Conseil général de l'économie, Inspection générale de l'administration, Contrôle général économique et financier, janvier 2019
- Rapport de la [Mission d'étude sur l'avenir des corps d'ingénieurs](#), Daniel Canepa–Jean-Martin Folz, janvier 2009

Annexe 4 : Méthodologie du traitement des offres de la place de l'Emploi Public

Les offres d'emploi publiées sur la Place de l'Emploi Public ne sont pas référencées dans les catégories de métiers définies par la mission. Des analyses et expérimentations ont été réalisées pour établir la catégorie du poste correspondant en exploitant le titre et le texte de l'offre lorsque celui-ci est disponible, c'est-à-dire lorsque l'offre n'est pas diffusée en format word et pdf dans une pièce que la mission n'a pas pu récupérer.

Voici les critères pour identifier l'appartenance d'une offre à l'un des métiers retenus par la mission :

- *data engineer* : le titre de l'offre contient ou bien « data engineer » ou bien « ingénieur donnée/data »
- *data architect* : le titre contient « architect » et une référence à la « donnée/data » ou au « cloud » (pour ne pas confondre avec les architectes applicatifs et architectes techniques)
- *data analyst* : le titre contient « analys » ainsi que « donnée/data »
- *statisticien* : le titre contient « statisti » ou le descriptif contient « statistic » (« statistique » est présent dans des descriptifs de nombreux postes qui ne sont pas des postes de statisticien)
- *data scientist* : le titre contient « scien » et « donnée/data »
- *expert décisionnel* : les mots « BI » ou « décisionnel » sont présents dans le titre ou dans le texte de l'offre.

Par ailleurs, on introduit dans la catégorie « data architect », les offres dont les descriptifs contiennent au moins deux des mots : hadoop, kafka, nosql, parallèle et datacenter.

De la même manière, on ajoute à la catégorie *data scientists*, les offres dont les descriptifs contiennent au moins deux des mots : python, scala, julia, spark ou deux des mots : intelligence artificielle, machine learning, apprentissage automatique, big data, fouille, datamin, massive.

Pour les statisticiens, on constate que nombre d'offres ne contiennent pas le terme « statistique » dans leur titre et sont présentées comme des postes de chargés d'études difficiles à discerner dans le temps dont dispose la mission, des chargés d'études qui ne sont pas statistiques. Les chiffres sont donc à prendre avec précaution.

Ces listes de mots ont été déterminées à la lecture des descriptifs de plusieurs offres mais ne peuvent prétendre être exhaustives. Un certain nombre d'offres n'ont pas de descriptif accessible et si cette part ou le champ général couvert par la Place de l'Emploi Public ont évolué sur les offres relatives aux experts de la donnée de la fonction publique d'État pendant la période 2016 – 2021, cela pourrait limiter la pertinence des résultats.

Sont comptabilisés par cette méthode, les offres et non les postes. Un poste sera comptabilisé autant de fois qu'un candidat est recherché. Un poste avec une forte rotation sera ainsi être comptabilisé plusieurs fois.

La catégorie de l'offre n'est pas remplie avant 2018 ce qui limite son usage. Par ailleurs, avec la méthode décrite ci-dessus visant à conduire à des métiers homogènes, on constate que les offres de *data scientist* sont réparties dans près de trente catégories différentes et celles de *data engineer* dans quinze. Même si l'évolution de la nomenclature fait accroître de façon factice ce nombre de catégories par métier, à la fois les limites de la nomenclature et de son appropriation par les employeurs rendent inutilisables ces catégories.

L'analyse de la description des offres montrent qu'à l'exception de quelques offres de *data analyst*, de statisticiens, et d'experts décisionnels pour lesquels il y a quelques postes de catégorie B, les postes sont des postes de catégorie A, et de catégorie A+ pour 15 % des offres de *data scientist*.

Annexe 5 : Questionnaires de la mission

1. Questionnaire envoyé aux ministères

Effectifs en experts de la donnée aujourd'hui dans votre direction par métier (en mars 2021 ou à défaut au 31.12.2020), effectifs en équivalent temps plein

<i>Data scientists</i>	
<i>Data engineers</i>	
<i>Data architects</i>	
<i>Data analysts</i>	
Polyvalents	
Autres experts de la donnée	Préciser métiers et compétences

Pour chacun des métiers, effectif total et effectif pour chacun des statuts suivants :

Fonctionnaires / CDI / CDD / Apprentis/ EIG/ Stagiaires / Autres statuts (préciser lesquels)

Si fonctionnaires, indiquer les principaux corps et leurs effectifs.

Appel à des compétences externes par métier (effectifs ETP à la même date)

Pour chacun des métiers en précisant leur nombre pour chacune des catégories suivantes :

Prestataires / Chercheurs / Autres externes (préciser lesquels)

Liste et effectifs des unités de la direction dans lesquelles il y a un ou plusieurs experts de la donnée

À quels besoins répondent ces experts ? Répondre par une intensité de besoin (4 très fort, 3 fort 2 moyen, 1 faible, 0 nul) et par une intensité en utilisation technique IA (même échelle).

Besoin	Intensité	Dont intensité en utilisation technique IA
analyse prédictive		
détection de valeurs atypiques		
retrouver une donnée particulière		
chatbot		
optimisation-automatisation d'un processus		
production de statistiques		
production de tableaux de bord		
cartographie		
constitution de lacs de données		
structuration de hub de données		
amélioration de la qualité des données		
codification automatique		
Autres : préciser		

Pour quelles finalités ? Mettre une intensité de besoin (4 très fort, 3 fort 2 moyen, 1 faible, 0 nul)

Finalité	Intensité
suivi et évaluation des politiques publiques	
modernisation et efficience de processus métiers	
améliorer le pilotage des politiques publiques	
optimisation des contrôles	
analyse de risque	
amélioration du service au public	
fiabilisation des données	
Autres : préciser	

Pour chaque profil, quelles filières de recrutement sont privilégiées ?

	Filières
<i>Data scientists</i>	
<i>Data engineers</i>	
<i>Data architects</i>	

Data analysts	
Polyvalents	
Autres experts de la donnée	

Besoin en experts de la donnée par métier au sein de votre direction (en effectifs)

	Aujourd'hui	Dans 2 ans	Dans 5 ans	À plus long terme
Data scientists				
Data engineers				
Data architects				
Data analysts				
Polyvalents				
Autres experts de la donnée				

... et à l'extérieur de votre direction (chercheurs, prestataires...)

	Aujourd'hui	Dans 2 ans	Dans 5 ans	À plus long terme
Data scientists				
Data engineers				
Data architects				
Data analysts				
Polyvalents				
Autres experts de la donnée				

Commentaires, informations supplémentaires pour aider à comprendre votre situation (nuance sur une réponse précédente, difficultés de recrutement pour certains profils...)?

Question posée au niveau ministériel :

Avez-vous une feuille de route ou un document qui explicite les enjeux en termes de données et les besoins à moyen terme ?

2. Questions supplémentaires posées aux SSM

Dans le cadre de votre mission de statistique publique, menez-vous aujourd'hui des travaux de *data science* (lesquels) ?

- À quels besoins répondent-ils ? Répondre par une intensité : 4 très fort, 3 fort 2 moyen, 1 faible, 0 nul

Hors mission statistique publique menez-vous aujourd'hui des travaux d'expertise de données (*data science* et autres) à destination d'autres entités du ministère (lesquelles et pour qui) ?

- À quels besoins répondent-ils ? Répondre par une intensité : 4 très fort, 3 fort 2 moyen, 1 faible, 0 nul

... et pour quelles finalités ? Répondre par une intensité de finalité : 4 très fort, 3 fort 2 moyen, 1 faible, 0 nul

De quels outils informatiques disposez-vous pour vos travaux en *data science* (logiciels et infrastructures spécialisés en *data science*) ?

Hors mission de statistique publique, avez-vous reçu des demandes de travaux sur des données (*data science*, statistique...) auxquelles vous n'avez pas pu répondre ? oui/non

- pour quels types de travaux ? – comment dans ce cas le besoin a-t-il été satisfait ?

Avez-vous connaissance de travaux sur des données (*data science*, statistique...) réalisés par d'autres directions du ministère qui auraient pu, selon vous, bénéficier de la prise en charge ou de l'accompagnement du SSM, y compris en apportant du conseil sur la façon de mener les travaux ?

- si oui, pour quels types de travaux ? – Pourquoi selon vous ces travaux sont réalisés en dehors du SSM ? – En quoi le projet aurait bénéficié du soutien du SSM ? Aurait-ce eu un impact majeur sur le projet ? – Savez-

vous quelles techniques sont alors mises en œuvre pour chaque projet ? – Pouvez-vous identifier les finalités de chaque projet ?

Pour quels types de travaux sur des données menés par les directions métiers estimez-vous que le SSM n'est pas fondé à intervenir ?

Pour l'avenir et sous réserve que les conditions nécessaires soient réunies :

- sur quels types de travaux et sous quelle forme (conseil méthodologique, réalisation...) le SSM pourrait-il selon vous intervenir pour des projets intéressant les directions métiers ?
- quelle adaptation des compétences et des moyens cela nécessiterait-il ?
- quelles autres conditions devraient le cas échéant être satisfaites ?

Plus généralement, quels sont pour vous les avantages comparatifs et les limites d'une intervention du SSM sur des projets en dehors de sa mission de statistique publique ?

Notamment : – avantages en termes de compétences ? – limites et difficultés éventuelles à concilier ce rôle avec celui d'un organisme de statistique ?

Quel est le positionnement actuel du SSM vis-à-vis de la fonction d'administration ministérielle des données et des pôles ministériels de data science ?

- Quel est le partage des rôles entre le SSM et l'AMD ? Comment le SSM est associé à l'action de l'AMD ?

- Est-ce que ce partage des rôles est :

formalisé

peu formalisé

pas du tout formalisé

n'existe pas, l'AMD est au sein du SSM

- quels sont les relations et le mode de collaboration avec les data scientists de la direction (ou du service) chargée du numérique et les pôles « data » au sein des directions métiers ?

- comment jugez-vous l'organisation actuelle ? Quelles évolutions vous paraîtraient favorables à une gouvernance et une exploitation plus efficiente des données au sein de votre ministère ?

Notamment du point de vue de la fonction d'AMD, de son profil, de son positionnement et des pôles « data » au sein du ministère.

- Quel rôle idéalement le SSM devrait-il jouer et pour quelles raisons ?

Annexe 6 : Liste des répondants à l'enquête

Ministères avec une réponse décomposée par direction/unité (et liste des unités ayant répondu)

Économie, des finances et de la relance : secrétariat général, CISIRH, DGFIP, DGDDI, DGE SDP3E, DGE SI Signaux Faibles, service à compétence nationale SISSE, service à compétence nationale Peren

Intérieur : SDITN, Direction du numérique

Solidarités et santé : DSS, direction Générale de la Santé (et direction du numérique commune au Travail, emploi, insertion)

Travail, emploi et insertion : direction du numérique, DGEFP, direction générale du travail

Transition écologique : DGAC, DGPR, direction des affaires maritimes (DAM), direction générale des transports, des infrastructures et de la mer hors DAM, CGDD / Ecolab, CGDD hors Ecolab, DRH/ Service Pilotage/DSNUMRH, SG/service du numérique, service du Haut Fonctionnaire de Défense et de Sécurité, DGPR/SRNV/SCHAPI

Ministères avec une réponse globale sans décomposition

Agriculture et alimentation

Armées (réponse centrée sur le secrétariat général, la mission l'a complétée avec des éléments fournis en entretien)

Éducation nationale, jeunesse et sports

Enseignement supérieur, recherche et innovation

Europe et affaires étrangères

Hors ministères : services du premier ministre, autorités indépendantes

DINUM, DILA, Cour des comptes, HCeres, Arcep, CRE,

Service statistique public

Insee, SSP (MAA), Dares (MTEI), DEPP (MENJS), DGCL (MI), SSM DGFIP (MEFR), Drees (MSS), SSM FP (MTFP), Jeunesse (INJEP), Justice (MJ), OED (MA), SDES (MTE), SIES (MESRI), SSMSI (MI)

Opérateurs

ANSM, ASP, Cerema, France Agrimer, Health data hub, Ifremer, IGN, Ined, IRSN, Météo France, OFB, Shom.

Annexe 7 : Modélisation du nombre de recrutements dans les années à venir

Afin de mesurer l'ampleur des recrutements et d'étayer sa recommandation de hausse des effectifs des corps de l'Insee, la mission a élaboré une maquette de calcul intégrant de nombreux paramètres pour évaluer différents scénarios dans les prochaines années.

Pour répondre au besoin des répondants à l'enquête en *data scientists* hors SSP (+ 100 postes tous statuts confondus d'ici 2023), la mission estime qu'il faudrait **augmenter les recrutements de fonctionnaires d'une quinzaine de spécialistes par an pendant plusieurs années**. Ce rythme ne suffirait pas à répondre en 2023 aux besoins à deux ans avec le seul levier des fonctionnaires mais permettrait de renforcer progressivement la proportion de fonctionnaires parmi les *data scientists*.

L'imprécision (quinzaine et non quinze) peut surprendre, elle n'est cependant que la traduction de l'impact de choix qui ne sont pas encore arrêtés et qui donnent lieu à un **certain nombre d'hypothèses** dont la maquette permet de mesurer les impacts :

- La première d'entre elles correspond à ce que les ressources humaines ne soient pas limitantes dans le développement de l'usage des données : l'ensemble des besoins exprimés par les employeurs doivent être pourvus et les plafonds d'emploi ou les réorganisations ne seront pas une limite à ces recrutements ;

- Une deuxième hypothèse est que le nombre de fonctionnaires qui sont aujourd'hui des *data scientists* (37) ne va pas baisser : ils vont ou bien rester *data scientist* ou bien être remplacés par d'autres *data scientists* « spontanés », qui, comme l'ont été les premiers, seront issus des corps des IPEF, de l'armement ou des administrateurs et attachés de l'Insee sans hausse de l'effectif de ces corps. La maquette prévoit donc le maintien de ce vivier.

Par ailleurs, pourvoir la centaine de postes de *data scientists* qui sont amenés à être créés dans les deux prochaines années ne nécessite pas d'augmentation du nombre de fonctionnaires si on prévoit de faire appel au seul vivier des agents contractuels. C'est seulement si on prévoit de maintenir la part de fonctionnaires observée actuellement (23 %) qu'il faudra former 27 fonctionnaires en deux ans. Et si on souhaite augmenter cette part de fonctionnaires pour la rapprocher à la fois de la part observée chez les opérateurs et la part identifiée par l'enquête pour les autres métiers de la mission ; c'est alors 98 fonctionnaires qui sont nécessaires pour faire en sorte qu'un *data scientist* de la fonction publique d'État sur deux soit fonctionnaire. Cette augmentation de la part de fonctionnaires, qui n'est pas liée aujourd'hui à une réflexion stratégique (cf partie 3.3.4), peut se faire à un rythme progressif et la hausse des recrutements être étalée dans le temps. Par exemple, dix promotions de dix personnes sur dix ans peuvent pourvoir les 98 postes supplémentaires aussi bien que cinq promotions de 20 personnes. Pour des raisons de démographie du corps et d'évolution des carrières, on privilégiera plutôt un effectif lissé dans le temps.

En pensant à moyen terme, le calibrage du nombre des recrutements doit permettre de faire face à l'augmentation des postes, et d'en pourvoir certains par des fonctionnaires, mais il faut également compenser les départs, du métier de *data scientist* ou de la fonction publique. En effet, même si la moitié des 270 postes (hors service statistique public) de *data scientists* étaient pourvus en 2023 par des fonctionnaires, il faudrait anticiper le renouvellement d'une partie des effectifs. Un taux de départ parmi les fonctionnaires de 3 % correspondant peu ou prou à une carrière de 40 ans et qui rejoint ce qui est actuellement observé dans les corps de l'Insee, suggère entre 3 et 4 recrutés par an nécessaires pour maintenir le niveau des effectifs. Si du fait des effectifs moins nombreux et de la relative jeunesse des recrutés le renouvellement ne sera pas le premier enjeu de la hausse des effectifs, il faut noter que le taux de 3 % peut paraître faible et un taux de 5 % – correspondant à une durée moyenne de vingt ans sur des postes de *data scientist* – paraît plus vraisemblable puisque, d'une part, une partie des *data scientists* seront attirés par le secteur privé ou du moins par d'autres postes que ceux de la fonction publique d'État, pour un temps ou définitivement, et que

d'autre part une autre partie va évoluer sur d'autres types de postes (reconversion ou évolution sur des postes de directions). Le taux de départ de 5 % porterait à 6 le nombre de recrutements de fonctionnaires-*data scientists* nécessaires pour maintenir les effectifs dans le temps.

En rassemblant ces éléments et cette hypothèse de 5 % de départs annuels, en intégrant quinze nouveaux fonctionnaires par an, il faudrait huit ans pour assurer la hausse et le maintien des effectifs associés à l'accroissement des besoins hors SSP attendus pour 2023 et pour accroître la part de fonctionnaires à 50 %. Cette durée peut paraître un peu longue et n'est donnée qu'à titre indicatif. Elle permet de ne pas avoir de rupture entre hausse des recrutements et renouvellement même s'il est toujours possible de baisser l'ampleur des recrutements quand arrive la phase de stabilisation des effectifs.

La quinzaine de recrutements ne cible que les effectifs hors service statistique public. Pourtant celui-ci exprime également des besoins en matière de *data science* et par suite c'est plutôt une vingtaine de recrutements de fonctionnaires qui seront nécessaires.

Outre le cas des *data scientists* pour lesquels la question de la part des fonctionnaires doit être traitée, **la maquette permet également de mesurer l'importance de la fidélisation sur le nombre de recrutements de l'ensemble des experts de la donnée.**

Même si les jeux d'hypothèses du tableau ci-dessous sont illustratifs ; ils permettent de montrer qu'en fonction d'une fidélisation améliorée ou non, le nombre de recrutements nécessaires peut varier dans de fortes proportions : 110 recrutements d'experts de la donnée par an pour maintenir les effectifs dans un cas ou bien 160 dans l'autre cas.

Deux scénarios de taux de renouvellement des experts de la donnée

Taux de départs	Fonctionnaires	CDI	CDD	Apprentis, EIG	Stagiaires	Autres statuts	Prestataires	Autres externes
Renouvellement bas	3 %	10 %	20 %	90 %	100 %	15 %	20 %	10 %
Renouvellement haut	5 %	20 %	40 %	90 %	100 %	25 %	30 %	20 %

Note : on a généralisé les cibles basses et hautes pour les fonctionnaires *data scientists* à l'ensemble des métiers

Dans ces derniers chiffrages est intégré le fait qu'une partie de l'augmentation en *data analysts*, celle associée au plan de la DGFIP fait d'ores-et-déjà l'objet d'un plan de formation auprès de fonctionnaires. Dans le scénario décrit plus haut, dans lequel la part de fonctionnaires augmente, et compte-tenu de leur plus grande stabilité, il y aurait entre 20 et 30 recrutements de moins par an.

Annexe 8 : Organisation des services statistiques ministériels

1. SSM directions d'administration centrale :

- SSM du ministère du Travail : direction de l'animation de la recherche, des études et des statistiques (Dares), rattachée au ministre
- SSM du ministère de l'Éducation nationale : direction de l'évaluation, de la prospective et de la performance (DEPP) au sein du secrétariat général
- SSM ministère des Solidarités et de la Santé : direction de la recherche, des études, de l'évaluation et des statistiques (Drees), rattachée au ministre

2. SSM services d'administration centrale :

- SSM ministère de la Transition écologique : service de la donnée et des études statistiques (SDES) du commissariat général au développement durable
- SSM ministère de l'Agriculture : service de la statistique et de la prospective (SSP) rattaché directement à la secrétaire générale

3. SSM sous-directions ou rattachés directement à un directeur d'administration centrale ou un chef de service

- SSM du ministère de la Justice : sous-direction de la statistique et des études, au sein du service de l'expertise et de la modernisation du secrétariat général.
- SSM sécurité intérieure : (SSM-SI) rattaché directement au directeur central de la police judiciaire
- SSM du ministère de l'Enseignement supérieur, de la Recherche et l'Innovation : sous-direction des systèmes d'information et des études statistiques (SIES) au sein du service de la coordination des stratégies de l'enseignement supérieur et de la recherche.
- SSM Fonction publique : sous-direction des études, des statistiques et des systèmes d'information au sein du service du pilotage des politiques de ressources humaines de la DGAFP.
- SSM du ministère de la Culture : département des études, de la prospective et des statistiques, au sein du service de la coordination des politiques culturelles et de l'innovation du secrétariat général.
- SSM Collectivités locales : département des études et des statistiques locales rattaché directement au directeur général des collectivités locales.
- SSM Immigration : département des statistiques, des études et de la documentation rattaché directement au Directeur général des étrangers en France
- SSM Douanes : département des statistiques et des études du commerce extérieur, rattaché directement à la direction générale.
- SSM études fiscales : département des études et statistiques fiscales, au sein du service de la gestion fiscale de la DGFIP

4. SSM bureaux et autres :

- SSM Défense : observatoire économique de la Défense, bureau rattaché à la direction des affaires financières au sein du secrétariat général pour l'administration
- SSM Jeunesse et sports ; mission enquêtes, données et études statistiques au sein de l'INJEP service à compétence nationale du ministère

Annexe 9 : Les compétences « données » des directions du numérique dans les ministères

- Ministère de l'Agriculture et de l'Alimentation

La délégation au numérique et à la donnée a pour rôle « de coordonner l'action du ministère pour accompagner le développement du numérique, que ce soit dans le fonctionnement et les outils de l'administration ou dans les politiques publiques dont le ministère à la charge ».

Le délégué au numérique et à la donnée est administrateur ministériel des données.

- Ministère des Armées :

Au sein de la DGNUM, la sous-direction de la transformation numérique :

« Assiste le directeur général dans l'exercice de ses responsabilités en qualité d'administrateur ministériel des données. À ce titre, elle :

- assure la maîtrise du patrimoine des données du ministère, en coordonnant l'action des états-majors, directions et services en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données et en définissant leurs responsabilités respectives
- veille à l'élaboration des outils, référentiels et méthodologies permettant une meilleure exploitation des données et un plus grand usage des sciences des données au sein du ministère ; »

- Ministère de l'Economie, des Finances et de la Relance

La délégation aux systèmes d'information : « veille à la coordination et à la convergence des systèmes numériques des ministères et impulse et met en œuvre une gouvernance ministérielle des systèmes numériques. Elle assure la fonction de délégué à la protection des données en lien avec la Commission nationale de l'informatique et des libertés.

À ce titre :

- elle met en œuvre et anime la fonction d'administration ministérielle des données... »

- Ministère de l'Éducation nationale, de la Jeunesse et des Sports

La direction du numérique pour l'éducation « Sans préjudice des compétences de la direction de l'évaluation, de la prospective et de la performance en matière de données statistiques, elle coordonne les actions en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données »

- Ministère de l'Europe et des Affaires étrangères

La direction du numérique « mène des expérimentations sur l'utilisation des données et l'intelligence artificielle pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers

(...) En liaison avec la direction interministérielle du numérique, elle coordonne les actions du ministère en matière d'informations publiques numériques et de données, de développement de services numériques et d'innovation numérique, de culture numérique ainsi que de gestion des compétences numériques. La direction du numérique assure la mission d'administrateur ministériel des données et coordonne l'action du ministère en matière d'inventaire, de gouvernance, de production, de large circulation et d'exploitation des données. »

- Ministère de l'Intérieur :

Au sein de la DNUM, la sous-direction de l'innovation et de la transformation numérique « assure un appui technique, logistique et financier aux services du ministère dans leur réflexion prospective sur la transformation numérique de leur métier...

... Elle assure le rôle d'administrateur ministériel des données et apporte son concours en tant que de besoin aux services du ministère dans la valorisation de leurs données.

Elle anime et coordonne les actions des services du ministère dans le cadre de la politique interministérielle d'ouverture des données publiques. »

Elle comprend un Laboratoire de la donnée et un laboratoire innovation

- Ministère de la Justice

« **Le service du numérique** définit, au regard des orientations interministérielles et en lien avec les directions et services, la stratégie de développement du numérique du ministère. Interlocuteur de la direction interministérielle du numérique, il organise et pilote les actions du ministère sur :...

- l'exploitation du potentiel offert par les données, en appui des politiques publiques du ministère ;... »

- Ministère des Solidarités et de la Santé :

Au sein de la DNUM, **la mission transformation numérique a notamment pour missions de :**
« Conseiller et accompagner la transformation numérique des ministères et de leurs directions, et identifier les opportunités de projets et de produits, animer la connaissance du patrimoine des données et valoriser le partage et l'usage des données »

La mission fabrique numérique, qui a pour missions de :

- »Valoriser le patrimoine des données des ministères en proposant des solutions innovantes d'exploitation de ces données. »

Ces missions doivent d'interpréter au regard de celles de la Drees (arrêté du 30 décembre 2019) :

Au sein de la Drees, la sous-direction de l'observation de la santé et de l'assurance maladie a pour mission :
« D'administrer les données de santé, c'est-à-dire de développer leur usage notamment à des fins de décisions opérationnelles éclairées, de définir les règles de leur mise à disposition à des fins de connaissance, d'étude, de recherche et d'évaluation, en liaison avec les producteurs de ces données et de favoriser leur accessibilité, tout en veillant à l'efficacité de leur processus de collecte »
Elle comprend un bureau « lab innovation et évaluation en santé »

Annexe 10 : Glossaire

AGD	Administrateur général des données
AMD	Administrateur ministériel des données
AMDAC	Administrateur ministériel des données, des algorithmes et des codes sources
ANSM	Agence nationale de sécurité du médicament et des produits de santé
Apec	Association pour l'emploi des cadres
API	Interface de programmation applicative
APIE	Mission d'appui au patrimoine immatériel de l'État
Arcep	Autorité de régulation des communications électroniques et des postes
ASP	Agence de services et de paiement
CGDD	Commissariat général au développement durable
CBCM	Contrôleur budgétaire et comptable ministériel
CGE	Conseil général de l'économie
CINUM	Comité interministériel du numérique
CITP	Comité interministériel de la transformation publique
CRE	Commission de régulation de l'énergie
DAM	Direction des affaires maritimes
Dares	Direction de l'animation de la recherche, des études et des statistiques (ministère du Travail)
DAE	Direction des achats de l'État
DEPP	Direction de l'évaluation, de la prospective et de la performance (ministère de l'Éducation nationale)
DGA	Direction générale de l'armement
DGAC	Direction générale de l'aviation civile
DGAFP	Direction générale de l'administration et de la fonction publique
DGE	Direction générale des entreprises
DGCCRF	Direction générale de la concurrence, de la consommation et de la répression des fraudes
DGDDI	Direction générale des douanes et des droits indirects
DGEFP	Direction générale à l'emploi et à la formation professionnelle
DGFIP	Direction générale des finances publiques
DGPR	Direction générale de la prévention des risques
DILA	Direction de l'information légale et administrative
DINUM	Direction interministérielle du numérique
DITP	Direction interministérielle de la transformation publique
Drees	Direction de la recherche, des études, de l'évaluation et des statistiques
DREETS	Direction régionale de l'économie, de l'emploi, du travail et des solidarités
DSI	Direction des systèmes d'information

DSS	Direction de la sécurité sociale
EIG	Entrepreneur d'intérêt général
Ensaе	École nationale de la statistique et de l'administration économique
Ensaі	École nationale de la statistique et de l'analyse de l'information
FTAP	Fonds pour la transformation de l'action publique
GENES	Groupe des écoles nationales d'économie et statistique
HCERES	Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur
IGN	Institut national de l'information géographique et forestière
IGPDE	Institut de la gestion publique et du développement économique
Ined	Institut national d'études démographiques
Inria	Institut national de recherche en sciences et technologies du numérique.
INS	Institut national de statistiques
Insee	Institut national de la statistique et des études économiques
IPEF	Ingénieur des ponts, des eaux et des forêts
IRSN	Institut de radioprotection et de sûreté nucléaire
ISIC	Ingénieur des systèmes d'information et de communication
ITPE	Ingénieur des travaux publics de l'État
NSIC	(filière) Numérique et systèmes d'information et de communication
OED	Observatoire économique de la Défense
OFB	Office français de la biodiversité
Peren	Pôle d'expertise de la régulation numérique
Rime	Répertoire interministériel des métiers de l'État
SDITN	Sous-direction innovation et transformation numérique (ministère de l'Intérieur)
SDES	Service de la donnée et des études statistiques (ministère de la Transition écologique)
SHOM	Service hydrographique et océanographique de la marine
SIES	Sous-direction des systèmes d'information et des études statistiques (ministère de l'Enseignement supérieur de la Recherche et de l'Innovation)
SSP	Service statistique public (regroupe l'Insee et les SSM)
SSP	Service de la statistique et de la prospective (ministère de l'Agriculture)

DINUM
Direction interministérielle
du numérique

20 avenue de Ségur - 75 007 Paris

 numerique.gouv.fr

  
@_DINUM

Insee
Institut national de la statistique
et des études économiques

88 avenue Verdier - 92 541 Montrouge

www.insee.fr

    
@InseeFr