
QU'EST-CE QU'UN·E CADRE DE LA STATISTIQUE PUBLIQUE ? ANALYSE HISTORIQUE DES FORMATIONS ET ANALYSE STATISTIQUE DES FICHES DE POSTE

Camille BEAUREPAIRE (), François COQUET (**), Ronan LE SAOUT (***)*

() Ministère de l'Économie et des Finances et Centre Maurice Halbwachs (CMH)*

*(**) Ensai et Irmair*

*(***) Ministère de la Transition Écologique et Crest-Ensai*

camille.beaurepaire@finances.gouv.fr, francois.coquet@ensai.fr, ronan.le-saout@ensai.fr

Mots-clés : Histoire de la Statistique, Emos, compétences

Domaine concerné : 12.1 Histoire, 19 Enseignement, éducation

Remerciements : Les auteurs remercient Laure De Maillard Taillefer pour l'extraction des données, ainsi que Gaël De Peretti et Pascale Piétri pour leurs précisions sur l'organisation et le contexte des campagnes de mobilité INSEE. Les auteurs remercient les participants aux JMS 2022 pour leurs remarques.

Résumé

La France présente la particularité au niveau international de former ses cadres de la statistique publique dans 2 écoles, l'Ensaï et l'Ensaë. Ces cadres bénéficient donc tous d'un corpus spécifique de formation d'une durée longue (2 ou 3 ans). Pour autant, le choix des deux écoles en termes de spécialisation est différent : la statistique publique est un débouché possible à l'Ensaë mais non une spécialisation (et le nombre d'élèves fonctionnaires représente moins de 10% des élèves), alors qu'elle fait l'objet d'une spécialisation distincte à l'Ensaï (labellisée Emos -*European Master of Official Statistics*) par Eurostat- concernant plus du tiers des élèves mais jusqu'à présent sans obligation statutaire de suivre cette troisième année de spécialisation. Un ou une cadre de la statistique publique ne se définit donc pas qu'à travers les enseignements suivis, puisqu'ils sont dans une large mesure (et ce de manière plus forte à l'Ensaë) communs avec ceux du cursus ingénieurs. L'expérience professionnelle et les postes occupés permettent de définir de manière complémentaire les compétences qui lui sont associées.

Cet article questionne le métier de cadre de la statistique publique à travers ce double prisme, la formation suivie et l'expérience professionnelle. En quoi consiste aujourd'hui le « métier de statisticien » (Volle, 1984) ? Se transforme-t-il au contact des nouvelles opportunités d'accès aux données (*big data*) et des nouvelles méthodes statistiques (*data science*) ?

Une première partie, historique, retrace en particulier l'histoire de la spécialisation en statistique publique à l'Ensaï, dans la continuité de l'article de 1995 d'Alain Desrosières qui étudiait dans le *Courrier des Statistiques* l'évolution des formations à l'Ensaë entre 1942 et 1996, date du déménagement en Bretagne de l'Ensaï, issue de la division CGSA de l'Ensaë. La formation statutaire des attachés.es de l'Insee est au niveau Bac+4 (alors que celle des administrateurs est au niveau Bac+5). Cette particularité a amené à la création d'une année de spécialisation non obligatoire avec des enseignements métiers autour de la statistique publique, qui s'est concrétisée par la création d'un master en statistique publique. Si des spécificités relatives à la statistique publique ont été pérennisées à travers ce master, depuis labellisé Emos, il existe une volonté de rapprocher (comme c'est déjà le cas à l'Ensaë) cette spécialisation du cursus ingénieur et d'y renforcer la place de la *Data Science*, enjeu fort des carrières dans la statistique publique pour les prochaines années.

Une deuxième partie analyse les postes proposés lors des campagnes de mobilité sur la période 2014-2021, et interroge une possible montée en puissance des métiers autour de la *Data Science*. À l'aide d'analyses textuelles ou de graphes de communautés, on identifie des communautés de compétences communes, mais pas d'essor notable de la *Data Science*. Ce constat n'est pas anormal, une formation académique vise en partie à anticiper des besoins futurs ou amenés à se développer sur un temps long. Il n'est donc pas possible de définir de manière unique ce qu'est un statisticien public ou une statisticienne publique. Ces cadres ont des connaissances et compétences spécifiques, fruits à la fois des formations suivies et de l'expérience professionnelle acquise. Ce qui les caractérise est finalement la diversité des métiers associés à la sphère de la statistique publique.

Abstract

In this paper, we try to give a characterisation of French official statisticians in terms of specific skills. We recall the long-term policy carried the French NSI, which provides its statisticians with a long and extensive training in two dedicated schools: Ensaë and Ensaï. We show how the Ensaï has dealt with the training of official statisticians since its creation and moving in the Ker Lann campus near Rennes, and how this training has evolved with the needs of statistical engineering, and lately the rise of « big data » and data science. In a second part, we analyze the skills mentioned in the profiles of positions opened yearly inside the French Public statistical system. Textual and graphical analysis allow us to build communities that are mostly, but not exactly, superposed to the families set a priori by the Insee. Additionally, we show that skills linked to data science, although increasing with years, remain quite rarely mentioned in the profiles. As a whole, we conclude that there is no straight way to define an official statistician in the French system through a set of specific skills.

1. Introduction

En 1995, Alain Desrosières étudiait dans le *Courrier des Statistiques* l'évolution des formations à l'Ensaë entre 1942 et 1996, date du déménagement en Bretagne de l'Ensaï, issue de la division CGSA de l'Ensaë. Il y décrivait notamment l'évolution, dans ce cadre, de l'identification des besoins de la statistique publique. La présente communication vise à étudier la formation des cadres de la statistique publique depuis cette date, et sa traduction en termes de compétences dans les postes occupés. Nous nous concentrons en particulier sur les formations de l'Ensaï, la statistique publique y restant un axe spécifique alors que cette thématique est moins distinguée au sein de l'Ensaë. Cette communication vise aussi à préciser ce qui caractérise les cadres de la statistique publique française, en s'interrogeant sur les compétences associées, obtenues dans le cadre des enseignements – la France ayant la particularité de former ces cadres spécifiquement aux métiers de la statistique publique - ou à travers les postes occupés.

Une première partie, historique, retrace l'évolution de la formation des cadres de la statistique publique à l'Ensaë (administrateurs·trices de l'INSEE) ou à l'Ensaï (attaché·e·s statisticien·ne·s de l'INSEE) au travers des principaux jalons qui l'ont façonnée. Concernant la formation de l'Ensaï, on y analyse les rôles qu'ont joués l'adossement de la formation des attaché·es à celle des ingénieur·es, la mise en place de la Formation Continue Diplômante des Attachés, la création du Master de Statistique publique, l'obtention du label Emos (*European Master of Official Statistics*) délivré par Eurostat, la généralisation de la formation à niveau Master, et enfin l'émergence de la Data Science (Donoho 2017) et sa traduction dans la statistique publique. On observe notamment la persistance d'une forte composante informatique affirmée dès l'origine comme un facteur de différenciation avec les enseignements dispensés à l'Ensaë. Par ailleurs, les enseignements spécifiques de la statistique publique ont connu un essor notable avec la création du Master de Statistique Publique, qui permet en outre de spécialiser les élèves en méthodologie statistique, études économiques ou informatique. On décrit enfin les plus récents développements de la formation visant à rapprocher le Master de la troisième année de formation d'ingénieurs de l'Ensaï. Cette partie s'appuie sur les rapports de l'Inspection Générale de l'Insee¹ et des documents pédagogiques internes à l'Ensaï. Les cadres formé·es à l'Ensaï et à l'Ensaë restent ainsi des « couteaux suisses » capables d'occuper un très vaste panorama de fonctions à l'issue de leur formation, panorama que nous proposons d'étudier en seconde partie.

Dans la deuxième partie, nous analysons donc les fiches de poste proposées dans le cadre des campagnes de mobilité, sur la période 2014-2021, en identifiant des communautés de métiers et de compétences, et en cherchant à les relier aux évolutions récentes des formations. Nous nous appuyons ainsi sur plusieurs millésimes de fiches d'offres de poste proposés dans le cadre des campagnes de mobilité. En plus de l'intitulé du poste, d'une description des missions et des caractéristiques du poste, chaque fiche détaille la liste des compétences associée au poste concerné. Cette liste est elle-même construite à partir d'une grille des métiers de l'Insee (2016). Une analyse textuelle nous permet d'identifier les termes clés associés aux métiers de la statistique publique (données, études, production...), sans identifier d'essor massif de la *Data Science*. Le nombre de postes incluant dans leur descriptif des termes relatifs à la Data Science est néanmoins en nette augmentation (de rien en 2014 à une quinzaine sur 700 en 2021). Des graphes de compétences (dans l'esprit de Jumah 2014, qui cherche à définir le métier de *Data Scientist* à partir de fiches LinkedIn) permettent d'identifier des communautés de métiers, attendues par l'Insee chez ses cadres, et de questionner la place des métiers mobilisant la *Data Science*.

¹ Ces rapports ne sont pas publics. Ils sont mentionnés mais non cités.

2. La formation à la statistique publique à l'Ensaï : 1996-2021

*« La ligne de l'Insee est claire : on veut des écoles ouvertes, dans lesquelles la majorité des étudiants soient des non-fonctionnaires (...), en cours de scolarité que les élèves fonctionnaires soient traités comme les autres, avec les mêmes choix. Il y a des raisons (...) liées à la mentalité que nous espérons ainsi favoriser chez les futurs fonctionnaires (...). D'autre part, on tenait beaucoup à ce que ces écoles soient ouvertes sur le monde extérieur, c'est-à-dire fassent appel de façon prioritaire à des enseignants venant du monde académique et extérieur, et venant aussi des milieux professionnels. »
Paul Champsaur, table ronde du 19 octobre à l'occasion des 10 ans de l'Ensaï.*

Cette profession de foi éclaire la logique dans laquelle l'Ensaï et l'Ensaë se sont construites pédagogiquement à la fin du vingtième siècle ainsi que, dans une certaine mesure, les évolutions qu'elles ont connues depuis. L'Ensaë a pleinement intégré les élèves fonctionnaires dans ses cursus, les métiers de la statistique publique devenant un débouché parmi d'autres et associé à quelques cours spécifiques au sein d'un cursus commun. La distinction de la formation entre élèves fonctionnaires et non fonctionnaires reste au contraire présente à l'Ensaï, en particulier lors de l'année de spécialisation. Mais cette distinction a aussi été associée à la construction pédagogique à la fois d'un Tronc Commun identique à celui des élèves non-fonctionnaires et d'une réelle spécialisation autour de la statistique publique, avec un nombre d'élèves fonctionnaires restant élevé (environ 50 chaque année). Et si les écoles continuent à se différencier ainsi que le notait Desrosières (1995), avec une orientation plus informatique de l'Ensaï et plus économique de l'Ensaë, elles se rapprochent également à travers des formations désormais construites autour de la *Data Science*. La suite de cette partie se concentre ainsi sur l'Ensaï, et détaille son histoire après son déménagement à Bruz, dans la continuité de l'histoire des écoles écrites par Desrosières (1995).

2.1. Première période : 1994-2012

2.1.1. L'adossement de la formation des attachés à une formation d'ingénieur

L'idée d'adosser la formation initiale des fonctionnaires de l'Insee à une école d'ingénieurs procède d'une logique assez visionnaire : garantir sur le long terme la qualité scientifique des cadres de la statistique publique en obligeant cette formation à suivre l'évolution de la demande en statistique (et plus largement en science des données) de diverses branches professionnelles au sein de la société civile, et non sur les seuls besoins de l'Insee.

Le passage de Philippe Tassi à la direction des études de la division CGSA de l'Ensaë a permis de prouver qu'une formation en statistique axée sur l'ingénierie et l'explosion des moyens informatiques disponibles trouvait des débouchés bien au-delà de la sphère statistique publique, soit en répondant à une demande croissante (la gestion des risques industriels et financiers, l'analyse statistique des données épidémiologiques ou d'essais cliniques), soit même en anticipant sur cette demande (le marketing quantitatif). C'est en partie ce constat qui a décidé l'Insee à transformer la « deuxième division » en une école autonome, l'Ensaï, dotée d'un personnel propre, puis de locaux propres (ceux qu'elle occupe toujours sur le campus de Ker Lann, près de Rennes). Cette école a été créée en 1994 avec l'objectif affiché de devenir une école d'ingénieurs pour ses élèves non-fonctionnaires, statut obtenu en 2005. La scolarité des élèves non-fonctionnaires est donc conçue depuis l'origine sur 3 ans après une classe préparatoire ou un autre cursus permettant l'accès en première année. Le refus de principe d'avoir une formation initiale de trois ans pour les futurs attachés (puis attachés-statisticiens) de l'Insee a donné lieu à un compromis limitant la scolarité des

élèves fonctionnaires à 2 années, suivies sur la base du volontariat par une Formation Continue Diplômante des Attachés (FCDA) à l'issue duquel était délivré le diplôme de l'Ensay. Par « diplôme de l'Ensay », il faut entendre le diplôme créé avec l'école (en fait un certificat portant le sceau de l'Insee), qui sera remplacé pour les non-fonctionnaires par le diplôme d'ingénieur dès que l'Ensay sera habilitée à le délivrer.

2.1.2. Une formation calée sur les deux premières années du cycle ingénieur

Pendant les quinze premières années de fonctionnement de l'Ensay, la scolarité des élèves fonctionnaires a donc été principalement calquée sur les deux premières années de la formation de leurs camarades non fonctionnaires, avec l'acquisition des bases fondamentales puis des outils professionnels nécessaires à l'exercice du métier à un niveau relevant de l'ingénierie : en première année, bases fondamentales des probabilités et des statistiques inférentielles, de l'économie et de la programmation informatique ainsi qu'un premier projet statistique, souvent basé sur des données issues de la sphère statistique publique ; en deuxième année, techniques statistiques multivariées (régressions, séries temporelles, analyses factorielles...), sondages, économétrie, et deux gros projets, l'un en informatique et l'autre en statistique, ce dernier visant à affronter une problématique relativement complexe, issue du monde extérieur, avec des données de qualité variable mises à disposition par les tuteurs de projets, et un véritable enjeu professionnel pour ces tuteurs.

En parfaite logique avec la ligne rappelée par Paul Champsaur dans sa déclaration de 2006, durant ces deux années, la part des enseignements spécifiques à la statistique publique était très restreinte, voire marginale. Il existait un stage de découverte du service statistique public (SSP) en fin de première année (à partir de 2003) ; quelques options pouvaient différer selon le statut des élèves (démographie vs finance d'entreprise, par exemple) ; le cours de données manquantes (ingénieurs) était mis en parallèle d'un cours de traitement de la non-réponse (fonctionnaires). L'essentiel des enseignements spécifiques aux fonctionnaires était concentré sur le mois de juin de leur deuxième année, juste avant leur première prise de poste, et alors que leurs camarades non-fonctionnaires étaient partis faire leur stage d'application statistique. Il faut toutefois signaler que cette similarité a connu des évolutions non négligeables, notamment lors de l'introduction de larges choix d'options au second semestre de la deuxième année, ces choix étant souvent dans les faits assez restreints pour les fonctionnaires.

2.1.3. La formation continue diplômante des attachés : un bilan mitigé

La FCDA, qui était censée faire office de troisième année de formation, se déclinait en 18 modules d'une semaine à choisir dans un catalogue abordant des thématiques extrêmement variées, et un mémoire professionnel d'assez haut niveau. Elle pouvait s'effectuer en un maximum de 6 années, et n'était ouverte aux attachés qu'après une année de présence sur leur premier poste. Cette formation n'a que partiellement joué le rôle attendu. On peut considérer qu'elle a eu un véritable intérêt sur le plan de la formation continue en ce que la grande majorité des anciens élèves fonctionnaires s'y inscrivaient et en suivaient quelques modules. En revanche, l'aspect diplômant a été un échec : d'une part seul un très petit nombre (moins de 10% d'une promotion en général) est allé au bout de la formation, la réalisation du mémoire professionnel notamment se révélant très dissuasive ; et d'autre part, le « diplôme » délivré n'avait guère d'autre légitimité que celle que lui conférait l'Insee : il n'était même pas inscrit au répertoire national des certifications professionnelles !

2.1.4. Un cas particulier : les « analystes »

Il est à noter ici qu'une catégorie spécifique d'élèves fonctionnaires a bénéficié très tôt d'une véritable troisième année : il s'agissait des élèves dont le premier poste se trouvait à la Direction du Système d'Information (DSI) de l'Insee. Pour des raisons historiques, les fonctionnaires ayant un poste dans ce service pouvaient bénéficier d'une prime d'analyste développeur pour peu qu'ils aient passé avec succès l'examen professionnel correspondant. La DSI a très rapidement demandé que la formation à cet examen spécifique (organisé au niveau du Ministère de l'Economie et des Finances) soit effectuée à l'Ensaï, et accompagnée d'une formation complémentaire très orientée vers l'informatique afin que les personnes concernées aient les compétences nécessaires à la spécificité de leurs postes. Ces élèves passaient donc, à la demande expresse de leur hiérarchie, une année supplémentaire à l'Ensaï. En-dehors de la préparation spécifique à l'examen professionnel d'analyste développeur, les enseignements prodigués étaient plus ou moins mutualisés avec la filière de troisième année spécialisée en informatique, sous des dénominations variables, pour les élèves ingénieurs. Cette exception n'a pas fait jurisprudence, et est restée exclusive de la sphère informatique de l'Insee. La demande récurrente d'une vraie troisième année de formation, formulée par les élèves fonctionnaires (notamment lors des Conseils de Perfectionnement de l'Ensaï), a été ignorée pendant quinze ans, et ce n'est qu'au tournant de l'année 2010 que la donne a considérablement changé.

2.2. La possibilité d'un diplôme de Master

Si, statutairement, un attaché-statisticien est un cadre formé au niveau Bac+4 dont les deux dernières années à l'Ensaï, si cette formation bénéficie d'une réelle reconnaissance et valorisation au sein de l'Insee et plus largement du SSP, elle ne s'accompagne d'aucun vrai diplôme (l'attestation de réussite délivrée par l'Ensaï depuis quelques années à l'issue du cursus en deux ans n'a pas plus de valeur que n'en avait le diplôme de FCDA). L'élève fonctionnaire sortant de l'école au bout de deux ans après être passé-e par les CPGE n'a potentiellement pas d'autre diplôme universitaire que son baccalauréat.

C'est peut-être parce qu'il n'était pas issu du « sérail » que Jean-Philippe Cotis a été le premier Directeur Général de l'Insee à ressentir cet état de fait comme une anomalie. Or, dès la fin de la première décennie du millénaire, il était patent que les formations à Bac+4, qui correspondaient à l'ancien niveau « Maîtrise » des cursus universitaires, devenaient obsolètes dans le contexte 3-5-8, autrement nommé LMD. Jean-Philippe Cotis a donc, lors d'un Conseil de Perfectionnement de l'Ensaï où la question de la scolarité en 3 ans avait été posée une nouvelle fois, ouvert la possibilité de créer une formation de Master permettant de doter les attachés-statisticiens d'un diplôme reconnaissant officiellement le niveau de formation qui était le leur, celui de l'ingénierie.

Par un heureux concours de circonstances, cette ouverture s'est faite au moment où l'Université de Rennes 1 composait son offre de formation pour 2012-2016. Grâce notamment au soutien actif d'Isabelle Cadoret, alors doyenne de la Faculté de Sciences Economiques de l'Université de Rennes 1, l'Ensaï a pu insérer au sein de la mention de Master de Statistique – Econométrie une spécialité « Statistique Publique » en co-habilitation. Cette proposition a reçu un avis très positif lors de son évaluation par l'AERES en 2010, et techniquement, il a été possible d'accueillir la première promotion du Master de Statistique Publique en septembre 2012, pour des premiers diplômes délivrés en 2013 (il est à noter que le terme « Master » a désigné exclusivement ce qui en est techniquement la dernière année, celle qui suit les deux années de formation à l'Ensaï). Trois

parcours furent créés, représentatifs des compétences des élèves de l'Ensaï : « Méthodologie de la Statistique Publique » pour la statistique, « Statistique et Traitement de Données » pour l'informatique, et « Etudes statistiques » pour l'économie et l'économétrie. Ce dernier parcours, ouvert un an après les deux autres, avait la particularité d'accueillir également des étudiant.es universitaires, et ses enseignements étaient principalement délivrés par et à la Faculté de Sciences Économique de l'Université de Rennes 1.

Il serait exagéré de prétendre que l'arrivée somme toute très rapide de ce Master dans le paysage a spontanément suscité l'enthousiasme au sein des services de l'Insee. Un rapport de l'Inspection Générale de l'Insee de septembre 2011 souligne que l'Administration (les employeurs) n'était pas « demandeuse » de cette évolution, mais il était difficile de revenir en arrière, d'autant que la différence de cursus et de diplôme rendait le parcours attaché de moins en moins attractif pour les lauréat.es des concours d'entrée à l'Ensaï. Il fut donc décidé d'entrer dans le nouveau système à titre expérimental pour 3 ans, sur la base du volontariat. Le Master remplacerait donc la FCDA (sous l'appellation « formation continue décalée »), 5 élèves pourraient suivre l'entièreté de la formation directement à l'issue de leur scolarité nominale (on parlait de « formation continue intégrée »), et la formation spécifique des analystes serait intégrée à ce Master. En outre, des mesures d'équivalence permettraient aux attachés en cours de FCDA de « basculer » vers le Master. Après plusieurs hésitations, il fut décidé également que les 5 élèves « en formation continue intégrée » auraient leur affectation sur leur poste en même temps que leurs camarades de promotion, mais ne le rejoindraient qu'au premier mars de l'année suivante, à l'issue de la partie théorique de leur formation. Ce choix a constitué un handicap dans leur recherche de premier emploi : sur les deux premières années du dispositif, aucune de ces personnes à très fort potentiel n'a ainsi trouvé de poste à la Direction Générale de l'Insee. Enfin, le Master comprenait un stage long (20 semaines minimum), effectué sur le poste de rattachement de la personne concernée dans le but de valoriser une partie de son travail : il s'agissait là d'une innovation majeure par rapport au mémoire professionnel nécessaire à la validation de la FCDA, qui au contraire était en grande partie détaché du travail fait sur poste.

La valorisation du travail de poste au niveau d'un mémoire de Master a été un véritable succès. S'il n'était pas évident d'emblée qu'on trouverait pour chaque attaché-e la matière nécessaire sur l'un de ses deux premiers postes (pour respecter le délai global permis dans le cadre de la FCDA), les résultats ont été bien au-delà des attentes les plus optimistes. Non seulement toutes les personnes concernées ont réussi à définir, avec leur hiérarchie, un sujet de stage du niveau académique requis, le cas échéant après discussion avec les responsables académiques du Master, mais bon nombre de mémoires ont fait preuve d'une qualité tout à fait remarquable, et largement au niveau de ce qui était demandé pour les stages de fin de formation des élèves du cursus ingénieur. On peut même dire que cela a permis de revaloriser la perception que certain.es fonctionnaires pouvaient avoir de leur travail dans le SSP, et finalement de réévaluer positivement la formation d'attaché au sein de l'Ensaï avec des compétences acquises d'un niveau équivalent à celui du diplôme d'ingénieur.

Techniquement, le Master de Statistique Publique (c'est-à-dire, pour les élèves fonctionnaires, sa dernière année) se faisait donc selon un cadencement imposé par la continuation du système FCDA, c'est-à-dire par sessions successives, chacune durant généralement 3 jours avec 6 heures d'enseignement par jour. En revanche, les contenus évoluaient sensiblement : l'offre de cours était plus restreinte, avec des enseignements de spécialisation supposant acquis les enseignements du Tronc Commun des deux premières années. Il faut noter que le contenu des enseignements faisait l'objet d'une discussion annuelle avec la Division de la Méthodologie de l'Insee, afin de garantir le

meilleur équilibre possible entre les besoins professionnels du SSP et l'exigence académique inhérente au diplôme décerné.

Le succès du dispositif a été bien au-delà des espérances puisque la très grande majorité des élèves fonctionnaires a non seulement opté pour cette nouvelle formation mais, grande différence avec l'ancien système, est allée jusqu'au bout de ses exigences, et environ 80% des attachés entrés à l'Ensaï après 2010, y compris les élèves recrutés par la voie du concours interne, sont *in fine* titulaires de ce Master.

Parallèlement à la montée en puissance de ce nouveau dispositif, un autre chantier important, au niveau européen mais auquel l'Insee, en la personne notamment d'Alain Trognon avait pris une part essentielle, arrivait à maturité : celui du Master Européen de Statistique Officielle (Emos : *European Master in Official Statistics*). Emos est un label créé par Eurostat pour favoriser l'émergence d'une communauté pédagogique autour de la statistique publique. Le Master « Statistique Publique » fut (et reste) le seul master français labellisé Emos parmi une trentaine en Europe. Même si le label n'a que partiellement rempli ses objectifs au niveau européen (le nombre d'étudiants inscrits dans les autres masters en Europe reste faible : les élèves de l'Ensaï représentent plus de la moitié des étudiants européens au sein d'Emos, du fait de l'association du recrutement avec une période de formation), il a fortement favorisé le développement des relations avec des partenaires européens, à travers par exemple la mise en place d'un double-diplôme avec l'Université *la Sapienza* (Rome) en 2019 ou l'engagement d'enseignants internationaux. Un enseignant-chercheur de l'Ensaï (et co-auteur de cet article) est membre du *board* du label, portant ainsi les particularités de l'Ensaï au niveau européen. Cette labellisation participe aussi à l'image internationale de l'Ensaï et est source de valorisation des travaux de ses élèves, notamment par le biais du concours du meilleur mémoire EMOS, donnant lieu à des exposés ou des posters lors des congrès européens NTTS -*New Technics and Technologies for Statistics*- organisés par Eurostat.

2.3. Vers une formation généralisée en 3 ans

Depuis 2016, et face au succès rencontré par la mise en œuvre de ce master de « statistique publique », plusieurs rapports de l'Inspection Générale de l'Insee (2016, 2017, 2020) en ont tiré des enseignements, ont discuté ce que devait être la formation des attachés de l'Insee dans le futur et comment pérenniser l'attractivité du concours. Les rapports de 2016 et 2017 tiraient un bilan positif du master et recommandaient de le renforcer, ainsi que de favoriser l'international et les « *soft skills* ». Suite à ces rapports, le nombre de places en formation intégrée fut augmenté à 10 (au lieu de 5), avec une campagne de mobilité désormais effectuée en 3^e année pour les élèves effectuant le master. Les possibilités de scolarité Erasmus et de scolarité de Master au sein d'un autre master Emos en Europe furent ouvertes aux élèves fonctionnaires lors de l'année scolaire 2018-2019. Les rapports de 2020 recommandent de généraliser le Master à l'ensemble des élèves, avec à terme le titre d'ingénieur associé à cette dernière année de spécialisation, notamment pour pérenniser l'attractivité des concours. Des besoins de compétences spécifiques et de niveau master, en particulier en informatique, sont également soulignés. Le Comité de Direction de l'Insee a ainsi décidé en 2020 de généraliser dans les années futures cette 3^e année aux élèves attachés.

En parallèle, les programmes de formation de l'Ensaï ont évolué, tant au sein du master que du cycle ingénieur. Le master « Statistique Publique » a été renommé en 2017 « Évaluation et Décisions Publiques », plaçant l'évaluation des politiques publiques au cœur de sa formation. La réflexion autour de la nouvelle accréditation 2022 est en cours et s'appuie sur la réflexion engagée sur le cycle ingénieur autour de la « *Data Science* ».

En effet, le métier de scientifique de la donnée a évolué ces dernières années avec l'émergence des concepts autour de la *Data Science* et des *Big Data*, mais avec des compétences associées pouvant être principalement statistiques ou informatiques selon les cas. Afin de préparer la future accréditation et l'audit de la commission des titres d'ingénieur (CTI) en 2021, la Direction des études de l'Ensaï a donc engagé à partir de 2015 une démarche (dite de « modernisation des cursus ») de réflexion pédagogique visant à améliorer la formation de ces élèves aux métiers de la « *Data Science* » et son positionnement dans ce cadre.

Cette démarche a consisté à identifier les besoins de recrutement dans les entreprises et les organisations ainsi que les compétences requises pour répondre à ces besoins. Dans le cadre de cette réflexion, de nombreux éléments ont été collectés : avis d'experts, analyse d'autres formations, enquêtes auprès des employeurs et des élèves, groupe interne de travail. De ces éléments ont été déduits les enseignements nécessaires pour acquérir ces compétences tout au long des trois années de la formation à l'Ensaï. Le conseil d'école de novembre 2016 a validé une démarche en deux temps pour faire évoluer les cursus de l'Ensaï. Une « modernisation » des cursus du Tronc Commun (les deux premières années de formation) a d'abord été mise en œuvre, pour notamment approfondir les connaissances des étudiants en apprentissage statistique et *Big Data*, renforcer l'autonomie, le travail en groupe et les « *soft skills* » et mieux lier les enseignements d'informatique et de statistique (avec par exemple un projet informatique de première année renommé « traitement de données »). Une nouvelle offre autour des filières de spécialisation du cursus ingénieur a été définie dans un second temps, avec en particulier l'ajout d'enseignements de *Machine Learning* dans l'ensemble des filières, le slogan de l'école « *Become a skilled Data Scientist* » s'appliquant ainsi à l'ensemble des filières de spécialisation.

Les changements mis en œuvre dans le tronc commun ont par ailleurs permis de clarifier la place des enseignements d'informatique et d'économie à l'Ensaï, comme compétences associées essentielles garantes de l'identité de l'Ensaï mais toujours en lien avec les enseignements de statistique. L'Ensaï continue ainsi à accorder une place plus importante à l'informatique et moindre à l'économie comparativement à l'Ensaë.

Nous pouvons terminer cette partie en revenant sur la citation introductive de Paul Champsaur. Les choix effectués ces dernières années pour les élèves fonctionnaires ont visé à les rapprocher encore plus de la formation des élèves ingénieurs (« ce qui est bon pour les ingénieurs l'est aussi pour les attachés »), avec par ailleurs une cohérence entre les recommandations des rapports de l'Inspection Générale et les changements introduits : Tronc Commun identique intégrant l'apprentissage statistique et le *big data*, ouverture vers l'international, augmentation des stages. La réflexion en cours sur le nouveau programme du Master va également dans ce sens, avec des enseignements de *Machine Learning* intégrés au Tronc Commun similaires à ceux des élèves non-fonctionnaires. Il reste pour les années futures à généraliser la 3^e année à l'ensemble des élèves et y associer le titre d'ingénieur. Une réflexion pédagogique et stratégique sera certainement de nouveau nécessaire, au vu de la place que pourraient occuper les cadres de la statistique publique dans les activités numériques de l'Etat, avec potentiellement un rôle d'essaiage des compétences des statisticiens publics (Brunet et al., rapport Denum-Insee 2021). Un point particulier de la formation des élèves fonctionnaires reste néanmoins la présence d'une spécialisation orientée « statistique publique » en dernière année de formation, qui la distingue de l'Ensaë et lui donne aussi une place particulière en France et en Europe.

Il nous reste désormais à étudier comment ces compétences acquises au sein de la formation sont mobilisées au sein des postes occupés, et si des évolutions peuvent être perçues ces dernières années vers la *Data Science*.

3. Une analyse statistique des métiers de la statistique publique

3.1. Description des données

Les mobilités professionnelles des cadres de l’Insee (corps des attachés et des administrateurs) sont centralisées chaque année à travers une campagne de mobilité nationale, hors postes d’encadrement supérieur. Il est demandé aux services (Insee, Services Statistiques Ministériels -SSM- et Autres) en septembre d’identifier et de décrire les postes ouverts à la mobilité pour une prise de fonction en septembre de l’année suivante. La description du poste à pourvoir est effectué par le service recruteur, en tenant compte de l’organisation et des objectifs futurs du service. Cette description peut donc ne correspondre ni aux fonctions occupées par la personne titulaire du poste ni aux compétences mobilisées jusqu’ici. Une liste des postes est ouverte à la consultation mi-décembre. Il y a ensuite environ deux mois pour effectuer des entretiens, et permettre aux cadres de l’Insee de classer les postes sur lesquels ils et elles désirent candidater et aux services de classer les candidatures aux postes. L’affectation est ensuite effectuée par les services RH de l’Insee, en collaboration avec les services, selon un processus *ad hoc*. Les jeunes cadres de l’Insee ont par ailleurs une obligation de mobilité sur leurs deux premiers postes, qui ont en général une durée de trois ans.

Nos données correspondent aux descriptifs des postes, hors données nominatives, rédigés dans le cadre des campagnes de mobilité 2014 à 2021. La structure des données est la même pour l’ensemble des années et correspond aux champs des fiches de poste : intitulé du poste, descriptif des fonctions et des perspectives ultérieures, principales caractéristiques, ainsi que les champs issus d’une grille des métiers de la statistique publique (document interne, dont la dernière version date de 2016) qui identifie 38 métiers, regroupés en 6 familles professionnelles (production, études, action régionale, informatique, fonctions support, stratégie et pilotage). Ces métiers sont associés à un ensemble de savoirs, savoir-faire, et savoirs comportementaux (« *soft skills* »). Lors de la rédaction d’une fiche de poste, une liste des compétences attendues est proposée, en fonction du métier. Toutes les compétences n’ont pas à être sélectionnées, et d’autres peuvent être ajoutées. Le niveau d’acquisition (connaissances de base, maîtrise, expertise) doit également être précisé. Les compétences associées à la description du poste peuvent être rédigées de manière différente pour un même sens (par exemple « connaissance des concepts de l’IPC » et « connaître les concepts de l’IPC »). Une analyse manuelle a été effectuée pour regrouper ces compétences.

Les données étant issues d’une application RH de gestion des campagnes de mobilité (« GMC3 »), elles ont fait l’objet de retraitements. Jusqu’à la campagne de mobilité 2017, il y avait deux tours d’affectation, les postes non pourvus au premier tour étant proposés dans un second tour. Dans la base de données brute, les postes proposés à la fois au tour 1 et au tour 2 sont présents deux fois. De plus, après 2018, on constate également la présence de doublons. Les doublons ont donc été éliminés en identifiant les postes ayant un même descriptif dans un même service au sein de la même campagne de mobilité et en ne conservant que le descriptif le plus récent (*i.e.* pour un poste proposé à la fois au tour 1 et au tour 2, on ne conserve que le descriptif du poste proposé au tour 2, potentiellement modifié par le service recruteur). On élimine ainsi 1519 observations de notre échantillon initial de 7369 postes, soit environ 30% des postes jusqu’en 2017 et environ 6% après 2018. 152 de ces observations correspondent en fait à des postes différents mais avec un même descriptif (un service proposant deux postes identiques ou plus la même année, ce qui peut correspondre à des cas de ré-organisation et de postes ouverts pour des raisons administratives). Après élimination des doublons, notre échantillon comporte *in fine* 5850 postes répartis sur 8 campagnes de mobilité, de manière relativement équilibrée. La campagne 2018 inclut cependant moins de postes et correspond à la mise en œuvre d’une campagne de mobilité en un tour. Le but de

notre étude étant de caractériser l'évolution du descriptif des métiers, des descriptifs identiques sont conservés au sein de campagnes de mobilité différentes. On constate ainsi que les 5850 observations correspondent à 5217 descriptifs différents, avec 4711 descriptifs uniques, 416 présents 2 fois, 68 présents 3 fois, et 22 présents plus de 4 fois.

Tableau 1 : Nombre de postes (uniques) par campagne de mobilité

Campagne	N	%
2014	689	12%
2015	745	13%
2016	827	14%
2017	720	12%
2018	604	10%
2019	797	14%
2020	764	13%
2021	704	12%
Total	5850	

Source : Outil RH de l'Insee, Calculs des auteurs.

3.2. Analyse descriptive des compétences et caractéristiques des postes occupés

3.2.1. Métiers et compétences

Parmi les 5850 postes, il y a 2194 postes (37,5%) de niveau attaché, 2217 (37,9%) de niveau attaché confirmé et 1439 (24,6%) de niveau administrateur. 2940 postes (50,3%) sont situés en Ile-de-France. 1941 (33%) sont des postes à la Direction Générale de l'Insee, 2140 (37%) en Directions Régionales, 885 (15%) en SSM centraux, 462 (8%) en SSM régionaux et 422 (7%) hors Service Statistique Public –SSP- (DG Trésor, Banque de France...). En général, la durée minimale indiquée est de 3 ans, et la durée maximale de 5 ans.

Les tableaux ci-dessous précisent la répartition des familles et des 10 principaux métiers dans l'ensemble des campagnes et en 2014 et 2021. Pour environ 1650 postes, 2 familles et 2 métiers sont indiqués. Les distributions par famille sont stables. La famille « stratégie et pilotage » est sous-représentée (et en particulier les métiers liés au management stratégique), ces postes étant généralement affectés par d'autres canaux.

Tableau 2 : Famille des métiers occupés

	N Total	% Total	N 2014	% 2014	N 2021	% 2021
Action régionale	1102	14,7%	132	15,6%	144	15,4%
Analyse et études statistiques	2076	27,7%	256	30,3%	251	26,8%
Fonctions support	603	8,0%	60	7,1%	80	8,5%
Informatique	935	12,5%	115	13,6%	112	12,0%
Production Statistique	2322	31,0%	244	28,9%	289	30,8%
Stratégie et pilotage	453	6,0%	37	4,4%	61	6,5%
	7491		844		937	

Source : Outil RH de l'Insee, Calculs des auteurs.

Les métiers d'études statistiques (hors analyse conjoncturelle) représentent environ 25% des postes, la conjoncture 8 %, les enquêtes 8 % et plusieurs spécialisations (recensement, indices, méthodologie, comptes, fichiers administratifs) comptent pour entre 2 % et 5 %. La conception

d'opérations statistiques est le 3^e métier le plus mentionné. Les métiers informatique sont également très présents, le 5^e métier le plus mentionné étant celui de « concepteur-réalisateur d'applications ». On ne perçoit pas d'évolutions fortes entre 2014 et 2021.

Tableau 3 : Principaux métiers occupés

	N Total	% Total	N 2014	% 2014	N 2021	% 2021
Chargé d'études statistiques nationales	1078	14,4%	119	13,6%	132	14,0%
Responsable ou chargé d'études régionales	772	10,3%	78	8,9%	103	11,0%
Concepteur d'opérations statistiques	646	8,6%	78	8,9%	80	8,5%
Chargé de synthèses économiques et d'analyses conjoncturelles	631	8,4%	98	11,2%	80	8,5%
Concepteur-réalisateur d'applications	410	5,5%	46	5,3%	61	6,5%
Manager opérationnel	385	5,1%	36	4,1%	56	6,0%
Chargé d'enquêtes auprès des entreprises	377	5,0%	39	4,5%	40	4,3%
Chargé de l'exploitation de fichiers administratifs	365	4,9%	40	4,6%	47	5,0%
Méthodologue	335	4,5%	29	3,3%	46	4,9%
Analyste-Programmeur	207	2,8%	27	3,1%	36	3,8%
Chargé d'enquêtes auprès des ménages	200	2,7%	21	2,4%	27	2,9%
Chargé du développement et de la valorisation des ressources humaines	180	2,4%	19	2,2%	18	1,9%
Chargé de l'appui à l'action régionale	162	2,2%	23	2,6%	13	1,4%
Expert de domaine	158	2,1%	17	1,9%	31	3,3%
Chargé du recensement de la population	144	1,9%	9	1,0%	17	1,8%
Chargé de répertoires entreprises ou personnes physiques	138	1,8%	13	1,5%	14	1,5%
Chargé de la gestion administrative du personnel	137	1,8%	14	1,6%	15	1,6%
Chargé de comptes	135	1,8%	18	2,1%	19	2,0%
Chargé d'infrastructure et d'outils informatiques	134	1,8%	0	0,0%	17	1,8%
Chargé d'indices	110	1,5%	12	1,4%	9	1,0%
Chargé de la coordination statistique	98	1,3%	23	2,6%	3	0,3%

Source : Outil RH de l'Insee, Calculs des auteurs.

Les principales compétences attendues sur le poste sont listées ci-dessous, avec la liste des 20 principaux savoirs, savoir-faire et *soft skills*.

Tableau 4 : Principales compétences des métiers occupés

Savoir	nb	%	Savoir-Faire	nb	%	Soft Skill	nb	%
Méthodes d'analyse statistique	2182	9,7%	Conduire une analyse, problématiser et synthétiser	2088	9,6%	Travailler de façon collaborative	3987	16,8%
Sas	1569	6,9%	Analyser les demandes et apporter les réponses à des besoins spécifiques	1750	8,0%	S'organiser pour réaliser son travail de façon efficace	1524	6,4%
Méthodes de conduite et de pilotage de projets	1154	5,1%	Rédiger les résultats d'une analyse, un bilan, une synthèse	1189	5,4%	Autonomie	1466	6,2%
Applications informatiques métier	610	2,7%	Organiser et animer un travail collectif (partenaires internes et externes)	1020	4,7%	S'exprimer (à l'écrit et/ou à l'oral)	1355	5,7%
Econométrie	540	2,4%	Mettre en œuvre les techniques statistiques et économétriques adaptées	807	3,7%	écouter, comprendre et reformuler	1234	5,2%
Statistique descriptive	527	2,3%	Planifier des tâches	737	3,4%	Alerter et discerner les enjeux	1084	4,6%
Logiciels statistiques professionnels (« R », SAS...)	505	2,2%	Analyser des données chiffrées, faire des simulations et des projections	709	3,2%	Respecter les contraintes	1052	4,4%
Économie locale et nationale	503	2,2%	Mettre en œuvre les techniques rédactionnelles du domaine d'investigation	646	3,0%	Rigueur quant à la conformité des procédures utilisées	1001	4,2%
Sources disponibles pour la production statistique	476	2,1%	Détecter les anomalies et les situations à risque	615	2,8%	Gérer son temps et planifier celui des équipes	893	3,8%
Anglais	462	2,0%	Jouer un rôle de conseil et d'aide à la décision	615	2,8%	Initiative	813	3,4%
Techniques d'enquêtes	462	2,0%	Capitaliser, partager et diffuser les méthodes et les connaissances	451	2,1%	S'adapter	683	2,9%
Architecture fonctionnelle et applicative	456	2,0%	Mettre en œuvre des consignes, des procédures et des règles	429	2,0%	Pédagogie	504	2,1%
Méthodes, normes et outils de développement	452	2,0%	Piloter un projet	410	1,9%	Esprit de synthèse	456	1,9%
Micro et macro-économie	437	1,9%	Utiliser des outils ou des logiciels spécifiques à l'activité	410	1,9%	Sens du dialogue	440	1,9%
Statistiques d'entreprise	435	1,9%	Mettre en œuvre les techniques de communication	403	1,8%	Gérer les priorités et les urgences	415	1,8%
Statistiques démographiques et sociales	424	1,9%	Élaborer et conduire une exploitation statistique	390	1,8%	Argumenter	409	1,7%
Partenaires extérieurs	420	1,9%	Arbitrer entre besoins et moyens	356	1,6%	Rendre compte	346	1,5%
Organisation et rôle des acteurs	408	1,8%	Analyser et modéliser les besoins métiers et fonctionnels	355	1,6%	Réagir en temps réel, de façon maîtrisée et à bon escient	344	1,5%
Autres logiciels statistiques (hors Sas)	393	1,7%	Gérer un projet	330	1,5%	Discretion, respecter la confidentialité	340	1,4%
Comptabilité nationale	345	1,5%	Fomer et conseiller	315	1,4%	Sens du service rendu et de la continuité de service	329	1,4%

Source : Outil RH de l'Insee, Calculs des auteurs.

3.3. Graphe des compétences

3.3.1. Construction et analyse descriptive

L'objectif est ici de créer une mesure de similarité entre nos différentes compétences, à partir de leur coprésence dans les fiches de poste. Nous ne retenons que les savoirs et savoir-faire, en mettant de côté les savoirs comportementaux (qui ne permettent pas une discrimination très efficace entre les différents métiers de la statistique publique).

Pour chaque couple de compétences distinctes, nous comptons leur nombre de coprésences dans les fiches de poste, et construisons ainsi une matrice d'adjacence décrivant notre réseau. Ce dernier est non-dirigé, et pondéré par le nombre d'apparitions de chaque compétence dans l'ensemble des fiches de postes (en agrégeant ici les différentes années de campagne).

Le réseau ainsi construit sur l'ensemble des années est un réseau dense (voir graphique 4 pour une représentation): muni d'une unique composante géante, son diamètre est de 6 (signifiant que les deux compétences les plus éloignées sur le réseau ne le sont que de six autres compétences) et la distance moyenne entre deux compétences prises au hasard de 1,9. Sa transitivité est de 43% (deux compétences reliées à une troisième ont une forte probabilité d'être reliées entre elles également), et sa densité de 14% (14% des liens possibles entre compétences sont réalisés). Des éléments de théorie des graphes sont précisés en encadré 1.

Encadré 1 : Statistique des réseaux et communautés

L'analyse statistique des réseaux est un outil ancien qui connaît depuis plusieurs décennies des développements majeurs (Kolaczyk, Csárdi 2014), du fait des nouvelles capacités informatiques et des possibilités d'application dans de multiples disciplines scientifiques : biologie, physique, épidémiologie, sociologie, sciences cognitives, sciences de l'information et étude des réseaux sociaux...

Formellement, un réseau est décrit par un ensemble de nœuds et un ensemble d'arêtes les reliant entre eux (s'inspirant en cela de la théorie des graphes). Des variables supplémentaires peuvent être fournies pour décrire les nœuds ou les arêtes. Dans cet article, le réseau étudié est non-dirigé et pondéré : cela signifie que les arêtes ne sont pas orientées (telles des flèches), et que le poids des arêtes n'est pas uniforme (plus des compétences sont reliées entre elles dans les données, plus l'arête correspondante aura un poids important). De nombreuses statistiques descriptives permettent de décrire la structure d'un réseau :

- Notre réseau est composé d'une unique composante centrale connexe. Toutes les compétences sont ainsi reliées les unes aux autres, fût-ce par une chaîne d'intermédiation, dont la longueur maximale sur notre réseau est 6 (on appelle cela le diamètre du réseau). La longueur moyenne d'un chemin entre deux compétences prises au hasard est de 1,9 ;

- La transitivité de notre réseau est de 0,43 – ce qui indique que si un nœud A est connecté à la fois à un nœud B et à un nœud C, alors la probabilité que B et C soient également connectés est de 43 % ;

- Pour chaque nœud, on peut calculer son « degré », c'est-à-dire le nombre d'autres nœuds auxquels il est connecté. Sur notre réseau, le degré des nœuds varie entre 2 et 309, avec une médiane à 47.

Pour simplifier l'information représentée sur un réseau, il peut être intéressant de rendre compte de "communautés", c'est-à-dire d'ensembles de nœuds ayant tendance à former plus de liens entre eux que vers l'extérieur. On parle de techniques de partition de graphe, ou bien de détection de communautés. Un grand nombre de méthodes ont été développées, nous ne citons plus bas que les

plus courantes : celles-ci peuvent s'inspirer de méthodes plus générales de classification (partition hiérarchique, fondée sur des scores *ad hoc* [*fastgreedy*] éventuellement construits par simulations [*walktrap*], ou bien sur les positions structurelles des nœuds [*edge betweenness*]), utiliser des résultats théoriques associés aux graphes (*partitionnement spectral*), ou atteindre une partition d'équilibre par simulation [*spinglass*, *label propagation*].

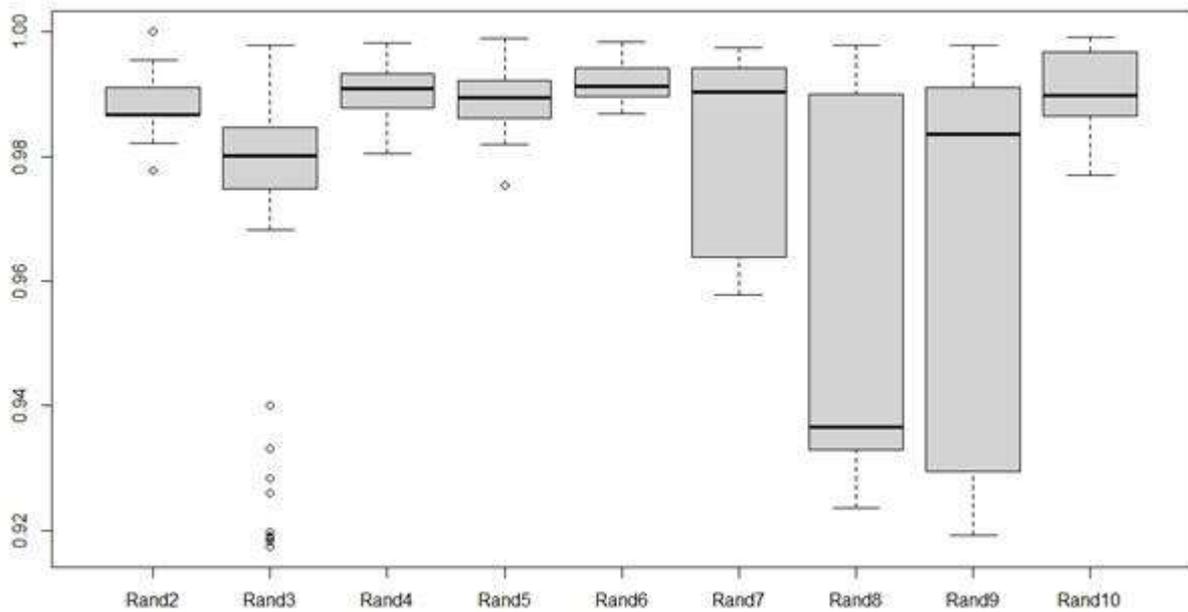
Ces algorithmes diffèrent dans leur fonctionnement, leur temps computationnel, leur caractère déterministe ou non ; mais aussi dans la définition implicite qu'ils retiennent pour définir les communautés. De ce fait, leur plus ou moins bonne performance dépendra en partie de la forme du réseau étudié en pratique, et des résultats attendus pour la recherche réalisée. De nombreuses comparaisons de performance ont déjà été menées entre ces différents algorithmes : elles reposent le plus souvent sur des simulations de réseaux artificiels (Yang et al., 2016; Orman, Labatut, 2009). Le choix de l'algorithme devrait dépendre de la taille du réseau (les méthodes les plus précises étant aussi les plus coûteuses en temps computationnel), et du *mixing parameter* correspondant à la proportion moyenne de liens avec des nœuds extérieurs à la communauté d'un nœud donné (sommée sur l'ensemble du réseau). Nous avons ici appliqué la méthode *spinglass*, réaliste au vu des propriétés structurales du réseau étudié (446 nœuds ; et l'algorithme *spinglass* est efficace quelle que soit la valeur du *mixing parameter*).

3.3.2. Communautés

Un algorithme de détection de communautés permet de synthétiser l'information contenue sur le réseau en partitionnant le graphe en un nombre fini de classes de compétences, homogènes du point de vue des liaisons sur le réseau. Deux classes d'une même communauté auront plus tendance à être reliées entre elles qu'avec des compétences d'une autre communauté.

Nous avons utilisé un algorithme commun de *spinglass*. Le résultat de l'algorithme dépend de son initialisation, aléatoire. Pour choisir le nombre de classes sur le réseau, on itère dix fois le partitionnement pour diverses valeurs du nombre de classes. Pour chaque paire de partitions obtenues conditionnellement à un nombre de classes, on calcule un indice de Rand. Les distributions des indices ainsi obtenus sont représentées dans le graphique 4 : plus la médiane tend vers 1, plus les partitions sont stables (et correspondent ainsi aux données). Ici, quel que soit le nombre de classes retenu entre 2 et 10, les indices de Rand sont en moyenne toujours supérieurs à 95%. On retiendra 6 classes de compétences, car les indices de Rand sont particulièrement hauts et concentrés autour de la médiane (mais c'est un choix en partie arbitraire).

Graphique 3 : Indices de Rand

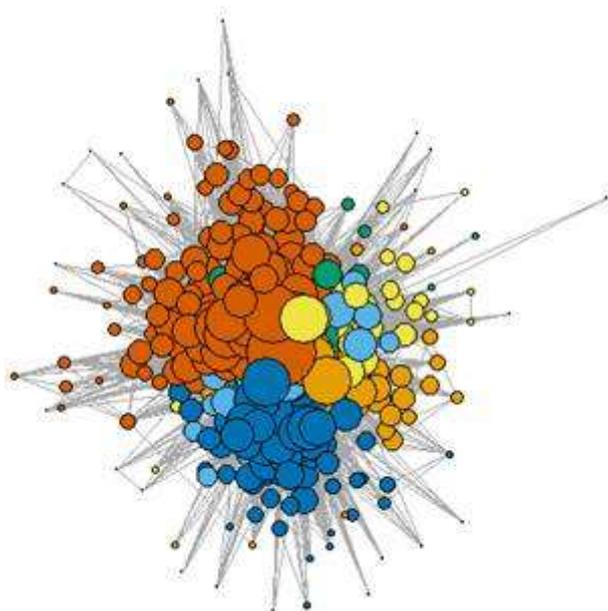


Source : Outil RH de l’Insee, Calculs des auteurs.

Note de lecture : pour les itérations de *spinglass* à 3 classes (« Rand3 »), la médiane des indices de Rand calculés deux à deux est de 0,98.

Les six communautés ainsi obtenues représentent différentes familles de métiers au sein de la statistique publique (graphique 5). Dans chaque communauté, nous distinguons des compétences plus ou moins centrales ou périphériques (tableau détaillé en annexe) : pour cela, nous utilisons la mesure de distance définie sur le réseau par le « plus court chemin », et calculons les indices silhouette de chaque compétence. Une compétence avec un fort indice silhouette sera dès lors centrale à sa communauté, indépendamment de sa fréquence d’apparition dans les fiches de poste.

Graphique 4 : Réseau des communautés



Source : Outil RH de l'Insee, Calculs des auteurs.

Note de lecture : Chaque cercle représente un nœud-compétence du réseau, chaque arrête un lien entre ces nœuds. La taille des nœuds est proportionnelle à leur nombre d'occurrences dans la base de données.

3.3.3. **Compétences et familles professionnelles**

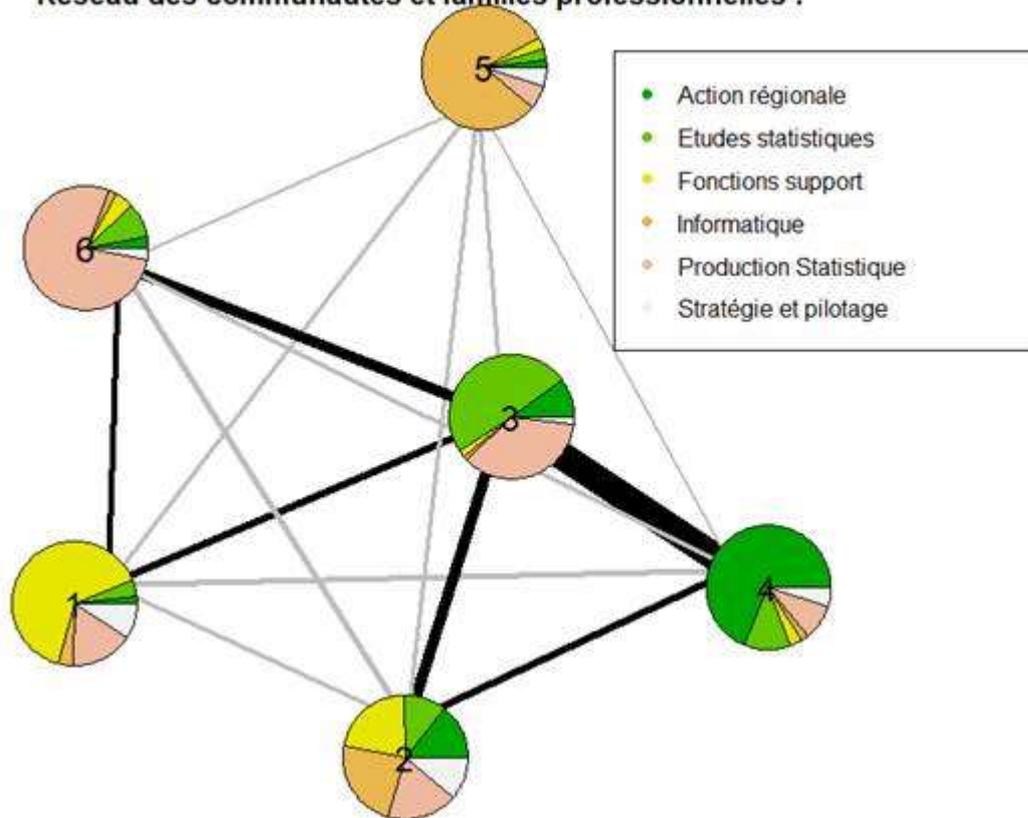
Chaque fiche de poste est caractérisée par une famille professionnelle (parmi six définies dans la grille des métiers de l'Insee) : nous pouvons dès lors regarder si telle ou telle compétence est reliée en pratique, *ex post*, à telle ou telle famille professionnelle. Pour cela, nous construisons le pourcentage d'apparition de telle ou telle famille professionnelle dans chacune des fiches citant telle ou telle compétence.

Pour chaque compétence, le maximum de ces pourcentages nous renseigne sur l'adéquation entre compétences et nomenclature des familles professionnelles. La médiane de ces maximums est de 80%, indiquant que plus de la moitié des compétences présentent un fort degré d'adéquation avec la grille des familles professionnelles. Mais le premier quartile est de 58% : un nombre non négligeable de compétences sont partagées par plusieurs familles professionnelles. C'est le cas de « savoir assurer une veille » (quasiment équiréparti entre les six familles), ainsi que « Produire des éléments d'aide à la décision » ou « Organiser et animer un travail collectif ».

L'association entre compétences et familles professionnelles permet de caractériser les communautés obtenues plus haut :

Graphique 5 : réseau des communautés et familles professionnelles

Réseau des communautés et familles professionnelles :



Source : Outil RH de l'Insee, Calculs des auteurs.

Note de lecture : Ce réseau est une représentation simplifiée des liens entre chaque communauté. La taille des arêtes inter-communautés est proportionnelle au poids des arêtes du réseau initial entre les nœuds composant chaque communauté (les plus importantes sont représentées en noir plutôt qu'en gris clair).

Un tel graphique rend compte de deux types d'informations : d'une part la plus ou moins bonne correspondance entre communautés et famille professionnelle, d'autre part le partage de compétences co-occurentes entre communautés.

Certaines communautés représentent ainsi fidèlement une famille professionnelle telle que définie par le secrétariat général de l'Insee : la communauté 5 renvoie aux compétences informatiques (dont on peut noter le relatif isolement sur le réseau en termes de spécificité des compétences), la communauté 1 aux fonctions support, la communauté 4 à l'action régionale, la communauté 6 à la production statistique.

L'étude des compétences fréquentes ou centrales à chacune de ces communautés (tableau 6) renforce ce constat. Ce point de vue complémentaire permet de préciser les compétences de la communauté 4 (en plus de l'action régionale, on trouve les compétences de publication et de diffusion), et de la communauté 6 (en plus de la production statistique, on trouve les compétences d'enquête, de collecte et de méthodologie).

Les communautés 2 et 3 sont moins facilement assimilables à une unique famille professionnelle. La communauté 3 renvoie spontanément aux études statistiques, ce que l'étude de

ses compétences fréquentes ou centrales (annexe) confirme. Mais certaines de ses compétences relèvent de la famille professionnelle de la production statistique et des enquêtes : cet alliage entre production et études semble constituer le cœur de métier de la statistique publique, du fait de la centralité de la communauté sur le réseau – c’est ici qu’on trouve les compétences les plus transversales et redéployables caractéristiques des métiers de la statistique publique.

La communauté 2 mélange à part quasi-égales les différentes familles professionnelles : l’étude de ses compétences fréquentes ou centrales révèle que l’on a ici affaire aux compétences d’encadrement, qui sont effectivement nécessaires dans chacune des familles professionnelles, tout en n’étant pas les plus transversales (en comparaison de la communauté 3).

On peut finalement nommer nos six communautés :

- 1 : ressources humaines et secrétariat général
- 2 : encadrement
- 3 : études statistiques
- 4 : action régionale, diffusion et publications
- 5 : informatique
- 6 : production statistique et enquêtes (méthodologie, collecte)

1.1.1. Caractérisation des communautés

En plus des familles professionnelles, les communautés peuvent être décrites par d’autres variables décrites sur les fiches de poste, comme la localisation institutionnelle du poste, ou encore le niveau attendu sur le poste (tableaux de pourcentages-lignes).

Les communautés sont inégalement réparties dans les différents environnements institutionnels du SSP : certaines compétences se retrouvent avant tout à la direction générale (DG) ou dans les directions régionales (DR) de l’Insee : ainsi des communautés 1 (ressources humaines), 2 (encadrement) et 5 (informatique).

Certaines communautés de compétences sont plus transversales : l’action régionale est (sans surprise) surreprésentée en DR et dans les antennes régionales des SSM, et la production statistique est également particulièrement présente en DR. Les études statistiques sont finalement la communauté de compétences qui irrigue au mieux les différents lieux d’exercice de la statistique publique (notamment « Autre »), ce qui confirme notre interprétation de cette communauté comme « cœur de métier ».

Tableau 6 : Communautés et appartenances institutionnelles

	Autre	DG	DR	SSM	SSM en région
Ressources humaines et secrétariat général	2	<u>48</u>	<u>46</u>	2	1
Encadrement	5	<u>45</u>	<u>35</u>	6	9
Études statistiques	<u>14</u>	<u>31</u>	<u>15</u>	<u>27</u>	<u>14</u>
Action régionale, publications et diffusion	4	<u>23</u>	<u>59</u>	4	<u>10</u>
Informatique	4	<u>45</u>	<u>44</u>	6	1
Production statistique et	2	<u>27</u>	<u>57</u>	13	2

enquêtes (méthodologie, collecte)					
---	--	--	--	--	--

Source : Outil RH de l’Insee, Calculs des auteurs.

Tableau 7 – Communautés et niveau d’exercice des postes

	Attaché	Attaché confirmé	Administrateur
Ressources humaines et secrétariat général	35	44	21
Encadrement	34	40	26
Etudes statistiques	44	29	28
Action régionale, publications et diffusion	45	40	15
Informatique	39	44	16
Production statistique et enquêtes (méthodologie, collecte)	39	46	15

Source : Outil RH de l’Insee, Calculs des auteurs.

1.1.2. Une montée en puissance des métiers de la *data science* ?

Le référentiel des métiers de l’Insee n’intègre aucune compétence renvoyant directement aux nouvelles techniques de la *data science*. Pour les approcher, nous regardons la présence de différents termes dans les descriptifs et intitulés de postes : « *big data* », « données massives », « apprentissage statistique », « *machine learning* », « *data science* », « science des données », « *data scientist* », « apprentissage automatique », « intelligence artificielle », « web scraping ».

Une seule fiche de poste comprenait l’un de ces termes en 2014, contre seize en 2021. Ces fiches restent cependant très minoritaires (2,3% des fiches en 2021), malgré leur progression importante.

Tableau 8 – Evolution de la fréquence des fiches « *data science* »

	2014	2015	2016	2017	2018	2019	2020	2021
Nombre de fiches « <i>data science</i> »	1	2	3	5	7	17	15	16
Pourcentage de fiches « <i>data science</i> »	0,1	0,3	0,4	0,7	1,2	2,1	2,0	2,3

Source : Outil RH de l’Insee, Calculs des auteurs.

Afin de caractériser plus avant ces fiches « *data science* », on peut se tourner vers les compétences qui les décrivent (tableau 9), et voir à quelles communautés du réseau elles appartiennent (tableau 10). Nous comptons ainsi le nombre de fois que chaque compétence apparaît dans ces fiches, et regardons quelles compétences sont surreprésentées dans cet échantillon « *data*

science » (par rapport à la base totale). Le dénominateur des pourcentages est ici le nombre total d'occurrences de compétences dans l'échantillon considéré.

Tableau 9 – Compétences caractéristiques des fiches *data science* (par ordre décroissant de surreprésentation)

Compétence	Pourcentage parmi les compétences de l'échantillon <i>data science</i>	Pourcentage parmi les compétences en population générale
Logiciels statistiques professionnels («R», SAS...)	6,0	1,1
Méthodes d'analyse statistique	9,3	4,9
Capitaliser, partager et diffuser les méthodes et les connaissances	4,5	1,0
Mettre en œuvre les techniques statistiques et économétriques adaptées	5,2	1,8
Rédiger les résultats d'une analyse, un bilan, une synthèse	5,4	2,7
Econométrie	3,7	1,2
Anglais	3,1	1,0

Source : Outil RH de l'Insee, Calculs des auteurs.

Tableau 10 – Communautés et *data science* (pourcentages à lire en colonne, rapportés au nombre de compétences)

Communautés	Echantillon « <i>data science</i> »	Population générale
Ressources humaines et secrétariat général	4,1	11,5
Encadrement	12,9	9,5
Études statistiques	51,6	34,5
Action régionale, publications et diffusion	7,1	17,3
Informatique	17,8	13,9
Production statistique et enquêtes (méthodologie, collecte)	6,5	13,3

Source : Outil RH de l'Insee, Calculs des auteurs.

Le tableau 10 permet de voir dans quelles communautés les compétences associées aux fiches « *data science* » sont surreprésentées : il y a une nette surreprésentation dans la communauté des études statistiques, ainsi que de plus légères surreprésentations en encadrement et en informatique. On retrouve ici plusieurs caractéristiques de ces nouveaux postes de *data science* : ils concernent plus souvent les postes de niveau administrateurs (plus souvent associés à des compétences d'encadrement), et sont orientés autant vers les études statistiques (mettant en œuvre des techniques innovantes) que vers la constitution de répertoires de données volumineuses (versant informatique du *big data*).

2. Conclusion

Le métier de "cadre de la statistique publique" n'a donc pas de caractérisation univoque en termes de compétences. Ce constat a déjà été celui des socio-historiens de la quantification, comme Volle (1984) qui distinguait la méthodologie, la technique et la publication au sein du « métier de statisticien » ; ou Desrosières (2013) qui distinguait « l'administrateur et le savant ». Tant par l'étude de la formation des élèves fonctionnaires que par celle de leurs fiches de postes, nous montrons la pluralité de leurs compétences. Leur formation reste dans une large mesure proche d'un cursus classique en statistique, et les évolutions passées et en cours renforceront cela. Un cadre de la statistique publique se caractérise en premier lieu par ses compétences statistiques ! Un label du type Emos permet néanmoins de valoriser les débouchés spécifiques à la statistique publique et de favoriser la réflexion académique sur les enseignements à y associer.

Les enjeux du *big data* et de la *data science* sont aujourd'hui bien identifiés par l'Insee (tant du point de vue de la formation, récemment modifiée en ce sens, que de celui des services recruteurs, qui proposent chaque année plus de postes liés à ces thématiques). Il manque à notre analyse une enquête auprès des élèves ou des statisticien.nes en poste pour compléter ce constat (Beaurepaire, 2018). La place de la *data science* à l'Insee reste aujourd'hui encore marginale. De nombreux facteurs d'inertie peuvent expliquer cette lente prise en compte (calendrier officiel resserré de publication, chaînes de production déjà écrites, personnels anciens et non formés à ces nouvelles techniques, resserrement continu des contraintes financières...). Certaines évolutions peuvent aussi être cachées (l'Insee a depuis toujours mobilisé des données massives) et certains outils se révèlent non nécessaires (les méthodes de machine learning sont complémentaires mais ne se substituent pas aux méthodes classiques). Les promesses de la *data science*, dont certaines restent à éclaircir, ne permettent sans doute pas de répondre à tous les enjeux de la statistique publique. Cette thématique n'en reste pas moins un enjeu futur qui demandera du temps pour essaimer (rapport Insee-Denum, 2021).

Bibliographie

- [1] Beaurepaire C. (2018), Projets professionnels et catégorisations ordinaires des attachés statisticiens stagiaires de l'INSEE, Mimeo.
- [2] Brunet F., Bourlange Danielle, Chignard Simon, Eidelman Alexis (2021), Rapport - Évaluation des besoins de l'État en compétences et expertises en matière de donnée, Direction interministérielle du numérique et Insee.
- [3] Desrosières A. (1995), D'une école de statistique et d'économie à l'ENSAE et l'ENSAI : 1942-1996, Courrier des statistiques, n°75-76, pp 47-53.
- [4] Desrosières, A. (2013). « L'administrateur et le savant : les métamorphoses du métier de statisticien » in. Gouverner par les nombres: L'argument statistique II. Presses des Mines
- [5] Donoho David (2017), 50 Years of Data Science, Journal of Computational and Graphical Statistics, 26:4, 745-766
- [6] Jumah Ferris (2014), *The Data Science Skills Network*, post Linked-In
- [7] Kolaczyk D., Csárdi G. (2014). *Statistical Analysis of Network Data with R*, Springer
- [8] Orman, G. K., & Labatut, V. (2009). A comparison of community detection algorithms on artificial networks. In International conference on discovery science (pp. 242-256). Springer, Berlin, Heidelberg.
- [9] Volle, M. (1984). Le métier de statisticien. *Economica*.
- [10] Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, 6(1), 1-18.

Annexe : Principales compétences des 6 communautés

Communauté	Nombre de compétences	Compétences parmi les plus fréquentes	Compétences les plus centrales (indice silhouette)
Communauté 1	63	Planifier des tâches Détecter les anomalies et les situations à risque Arbitrer entre besoins et moyens Dispositifs et outils « Ressources humaines » Textes administratifs et réglementaires du domaine	Insertion d'une personne en situation de handicap Appliquer les règles et les procédures de la commande publique Règles de la commande publique
Communauté 2	44	Organiser et animer un travail collectif Capitaliser, partager et diffuser les méthodes et les connaissances Former et conseiller Secret statistique Faire émerger les bonnes pratiques et en assurer la diffusion	Productions du Système Statistique Public (inventaire) Former et conseiller Méthodes, normes et outils d'organisation du travail Faire émerger les bonnes pratiques et en assurer la diffusion
Communauté 3	40	Méthodes d'analyse statistique Conduire une analyse, problématiser et synthétiser Sas Rédiger les résultats d'une analyse, un bilan, une synthèse Économétrie	Être capable d'effectuer une présentation synthétique Méthodes d'analyse statistique Sources disponibles pour la production statistique Accompagner les équipes dans les réalisations des tests
Communauté 4	49	Analyser les demandes et apporter les réponses à des besoins spécifiques Jouer un rôle de conseil et d'aide à la décision Économie locale et nationale Partenaires extérieurs Mettre en œuvre les techniques de communication	Pratique journalistique Publications nationales et régionales Choisir les modes de représentation adaptés aux messages et aux publics Assurer les fonctions de rédacteur en chef
Communauté 5	97	Architecture fonctionnelle et applicative Méthodes, normes et outils de développement Piloter un projet Analyser et modéliser les besoins métiers et fonctionnels	Outils de supervision, de prise en main à distance, de gestion de parc, de suivi des incidents Analyse conception orientée service Stratégie métier de l'entreprise et processus métiers
Communauté 6	153	Connaissance du domaine d'expertise (formation générale des enquêteurs, Capi ...) Théorie et/ou principe des indices Savoir gérer un calendrier et des échéances Nouvelles Conditions d'Emploi des Enquêteurs (NCEE) Techniques et méthodologies statistiques	Savoir évaluer les compétences des agents Connaître les concepts des indices de Prix Connaître le processus, les consignes et le calendrier opérationnel (note de cadrage) Organisation du site ou du pôle Savoir analyser les situations, apprécier la fiabilité de l'information reçue ou donnée