



Commission économique pour l'Europe

Conférence des statisticiens européens

Soixante-neuvième réunion plénière

Genève, 23-25 juin 2021

Point 7 c) de l'ordre du jour provisoire

Rapports, directives et recommandations élaborés sous les auspices de la Conférence :

Évaluer la qualité des sources administratives à utiliser dans les recensements

Directives pour l'évaluation de la qualité des sources administratives à utiliser dans les recensements

Note établie par l'Équipe spéciale de l'évaluation de la qualité des sources administratives à utiliser dans les recensements

Résumé

Le présent document est une version abrégée des « Directives pour l'évaluation de la qualité des sources administratives à utiliser dans les recensements ». L'objectif des Directives est de fournir aux producteurs de recensements un guide pratique pour évaluer la qualité des données administratives, au cours d'une série d'étapes d'évaluation : source, données, processus et résultats, la première de ces étapes étant le principal sujet de ces directives.

Les Directives ont été établies par l'Équipe spéciale de l'évaluation de la qualité des sources administratives à utiliser dans les recensements, dont les membres sont les suivants : Royaume-Uni (Président), Allemagne, Autriche, Canada, Espagne, États-Unis d'Amérique, France, Irlande, Israël, Italie, Monténégro, Nouvelle-Zélande, Pays-Bas, Pologne, Portugal, République de Moldova, Turquie, Fonds des Nations Unies pour la population, Eurostat et un expert indépendant.

La présente version abrégée des Directives a été préparée à des fins de traduction. Elle comprend la majeure partie de l'introduction, des chapitres sur le cadre de qualité et l'étape relative aux sources, ainsi que le chapitre complet des conclusions et recommandations. Veuillez noter que la numérotation des sections dans cette version abrégée ne correspond pas à la numérotation des chapitres et des sections dans les Directives complètes.

Les autres sections se trouvent toutes dans la version complète des Directives en anglais, qui a été envoyée pour consultation à tous les membres de la Conférence des statisticiens européens en mars-avril 2021 et qui est également disponible sur la page Web de la Conférence, à l'adresse suivante : <https://unece.org/statistics/events/CES2021>. Un résumé du résultat de la consultation sera présenté dans le document ECE/CES/2021/3/Add.1.

Sous réserve d'un résultat positif de la consultation, la session plénière de la Conférence sera invitée à approuver les Directives.



I. Introduction

A. Contexte

1. En 2017, l'Équipe spéciale de la CEE chargée des recensements fondés sur des registres ou mixtes a établi les Directives concernant l'utilisation de registres et de données administratives pour les recensements de la population et des habitations¹. Ces directives comprenaient une section sur les sources de données et leur qualité, accompagnée d'une discussion générale sur le sujet. Les experts de la Réunion d'experts CEE-Eurostat sur les recensements de la population et des habitations (Genève, 4-6 octobre 2017) ont désigné la qualité des sources administratives comme un sujet de première importance pour de nombreux pays. En conséquence, la Réunion d'experts a préconisé la création d'une nouvelle équipe spéciale de la CEE sur la mesure de la qualité des sources administratives à utiliser dans les recensements, qui s'appuierait sur les travaux de l'équipe spéciale précédente.

2. L'Équipe spéciale de l'évaluation de la qualité des sources administratives à utiliser dans les recensements a été créée en 2018, son mandat² ayant été approuvé lors de la réunion de février 2018 du Bureau de la Conférence des statisticiens européens, tenue à Helsinki (14-15 février 2018). Elle a fait rapport au Groupe directeur de la CEE sur les recensements de la population et des habitations, qui fait à son tour rapport à la Conférence et à son Bureau.

3. L'objectif de l'Équipe spéciale était de définir des orientations sur la mesure de la qualité des sources administratives à utiliser dans les recensements³. Le mandat stipulait que ces orientations devaient pouvoir être utiles à tous les pays de la CEE et qu'elles devaient s'appuyer sur les travaux du projet ESS.VIP ADMIN⁴ d'Eurostat sur l'utilisation des sources administratives dans la production de statistiques officielles.

4. Les présentes Directives constituent une boîte à outils pratique pour l'évaluation et la mesure de la qualité des sources administratives utilisées pour les recensements de la population et des habitations.

B. L'utilisation des données administratives dans les recensements

5. Les sources de données administratives sont des ensembles de données qui contiennent des informations recueillies essentiellement à des fins administratives. Il s'agit notamment de données recueillies par les services de l'État et les organismes publics et autres aux fins d'inscription de bénéficiaires, de réalisation d'opérations et de conservation des informations, habituellement dans le cadre de la fourniture d'un service. Ces ensembles de données comprennent les registres administratifs (avec identifiant unique) tels que les registres de la population, des entreprises, des adresses, de l'éducation, de la santé, de l'emploi et des impôts d'un pays, ainsi que d'autres sources administratives (sans identifiant unique). Les registres administratifs et/ou d'autres sources administratives sont utilisés pour créer des registres statistiques, qui sont spécifiquement utilisés à des fins statistiques, notamment pour le recensement.

6. L'utilisation de sources de données administratives dans les recensements varie selon les pays. Ces sources peuvent être utilisées pour améliorer ou compléter un recensement traditionnel, pour réaliser un recensement mixte ou dans le cadre d'un recensement entièrement fondé sur les registres. On observe une nette tendance à l'utilisation accrue des données administratives dans les recensements. Cette tendance a été motivée par les

¹ Disponible à l'adresse suivante : <http://www.unece.org/index.php?id=50794>

² Disponible à l'adresse suivante :

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2018/February/06Add.1-TF_on_quality_of_admin_data_for_censuses_ToR_apr.pdf.

³ L'Équipe spéciale a ensuite décidé d'adapter son titre et l'objectif correspondant en parlant d'« évaluation » plutôt que de « mesure » de la qualité des sources administratives à utiliser dans les recensements.

⁴ De plus amples informations sur ce projet sont disponibles à l'adresse suivante :

https://ec.europa.eu/Eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources_en.

avantages que les données administratives peuvent apporter, notamment la réduction des coûts et de la charge de travail des répondants, l'amélioration de la rapidité et de la fréquence des résultats, l'amélioration de la qualité et une plus grande souplesse dans la réponse aux besoins des utilisateurs (voir, par exemple, la section 4.1 de UNECE 2018). En outre, les conditions dans lesquelles l'utilisation de données administratives dans l'ensemble des systèmes statistiques nationaux est encouragée et facilitée ont changé dans de nombreux pays (voir la section 4.2 de UNECE 2018). Cela se traduit notamment par des changements dans la législation et dans l'acceptation par le public et les parties prenantes, ainsi que par l'évolution des techniques et des méthodes statistiques.

7. L'importance des données administratives a été mise en évidence par les difficultés que rencontrent aujourd'hui les organismes nationaux de statistique lorsqu'il s'agit de collecter des données directement auprès de la population, que ce soit en raison de la réticence du public à participer au recensement ou de leur capacité à le faire. Ce problème a été souligné au début de la crise de la Covid-19 en 2020, lorsque la capacité du public à communiquer avec les organismes nationaux de statistique et, parallèlement, la capacité de ces derniers à communiquer avec le public ont été considérablement amoindries.

C. Principaux risques pour la qualité

8. Malgré tous les avantages que peuvent apporter les données administratives, il existe un certain nombre de considérations essentielles en matière de qualité qui doivent être évaluées avant qu'une source administrative soit incorporée dans un recensement. Premièrement, l'organisme national de statistique n'aura qu'un contrôle limité sur la façon dont les données sont recueillies et traitées. Il existe donc une dépendance importante à l'égard des autorités qui détiennent les données administratives. Par exemple, si l'autorité administrative n'est pas en mesure de répondre aux exigences de l'organisme national de statistique en ce qui concerne la fourniture des bonnes données au bon moment, cela aura une incidence sur le degré d'actualité des résultats du recensement. De même, si l'autorité administrative ne communique pas dûment avec l'organisme national de statistique au sujet de tout changement potentiel de la source, cela pourrait influencer sur la cohérence et la comparabilité des données.

9. Deuxièmement, l'utilisation de données administratives par l'organisme national de statistique à des fins autres que celles pour lesquelles les données ont été initialement collectées soulève des problèmes de confidentialité, de sécurité et de droit. L'organisme national de statistique doit donc évaluer l'acceptabilité par le public, en veillant à ce que les garanties requises soient en place et que le public (et l'autorité administrative) en soient informés. L'utilisation doit également être légitime. En l'absence d'acceptation ou d'accord de la part du public et du fournisseur de données, ou d'une base juridique crédible pour l'utilisation de la source administrative, la réputation de l'organisme national de statistique et sa capacité de réaliser un recensement de haute qualité seraient fortement compromises. Cela peut se produire si le public modifie son comportement et sa façon d'interagir avec l'autorité administrative ou l'organisme national de statistique, en raison de préoccupations concernant la manière dont l'organisme national de statistique utilise ses données.

10. Troisièmement, les données administratives n'ont (généralement) pas été collectées à des fins statistiques. Par conséquent, les sources de données peuvent avoir adopté des concepts et des définitions différents de ceux que requiert le recensement ; elles peuvent se fonder sur des périodes de référence différentes ; et leur couverture de la population recensée peut être limitée. En outre, la précision et l'exhaustivité des données dépendront fortement de l'importance des données pour la fonction de l'autorité administrative. Les sources administratives peuvent également être sujettes à des changements dans le temps et à des incohérences dans la façon dont les données sont collectées dans les différents groupes de population.

11. Enfin, la complexité des données administratives ainsi que la disponibilité et l'exhaustivité des métadonnées associées auront une incidence sur la capacité d'un organisme national de statistique à comprendre, consulter et utiliser une source administrative. Par exemple, les données administratives peuvent être conservées dans des structures de données

vastes et complexes, ce qui soulève des difficultés techniques importantes que l'organisme national de statistique doit prendre en compte et surmonter. La complexité des données administratives peut également avoir une incidence sur l'accessibilité et la clarté des résultats du point de vue de l'utilisateur. En d'autres termes, les utilisateurs du recensement peuvent avoir du mal à comprendre l'utilisation de données administratives dans le recensement et les conséquences de cette utilisation sur la qualité des résultats du recensement.

12. Ces considérations essentielles en matière de qualité guideront les décisions relatives à l'utilisation de données administratives dans un recensement. Les Directives abordent chacune de ces considérations en détail.

II. Cadre de qualité

13. La qualité des statistiques dépend de l'adéquation de la production statistique à l'usage qu'il est prévu d'en faire. Par exemple, la définition de la qualité par le système statistique européen (SSE) est dérivée de la famille de normes ISO 9000 : mesure dans laquelle un ensemble de caractéristiques inhérentes à un objet répond aux exigences (ISO 2015). Dans les statistiques officielles, l'objet peut être un produit, un service, un processus, un système, une méthodologie, une organisation, une ressource ou une entrée [de données] statistiques (Eurostat 2020, p. 17). Dans le cadre d'un recensement, la qualité des données administratives utilisées doit donc être examinée en fonction de la manière dont ces données sont recueillies et traitées par les fournisseurs et les organismes nationaux de statistique, jusqu'aux résultats finaux du recensement.

A. La qualité et les erreurs dans les recensements

14. Lorsque les statistiques officielles sont produites à l'aide d'une méthode d'enquête par sondage, les questions de l'enquête sont conçues et testées de manière à réduire les erreurs de mesure et à garantir ainsi une précision et une fiabilité maximales. Par conséquent, l'erreur des estimations produites est supposée être causée par un échantillonnage déficient et est généralement mesurée et signalée au moyen du système de l'erreur quadratique moyenne et/ou des intervalles de confiance. Toutefois, ces mesures ne tiennent pas compte des erreurs non dues à l'échantillonnage, qui sont particulièrement importantes dans le contexte des recensements, où l'objectif est de prendre en compte l'ensemble de la population. Ainsi, comme pour les statistiques produites à partir de données administratives plus généralement, les principales sources d'erreur observées dans le contexte des recensements ne sont pas les erreurs d'échantillonnage, mais les erreurs de représentation (couverture) et de mesure (Zhang 2012). Une pratique courante consiste donc à ajuster les estimations du recensement sur la base des résultats d'une enquête postdénombrement – bien que cela puisse donner lieu à des controverses car les méthodes complexes de double estimation sont mal comprises par le public⁵.

15. Lorsque des données administratives et d'autres sources de données telles que les mégadonnées sont utilisées dans les recensements, l'éventail des erreurs possibles est plus large que dans un recensement traditionnel, car les processus de collecte de données ne sont pas contrôlés par les organismes nationaux de statistique. Zhang (2012), s'appuyant sur Groves *et al.* 2004, distingue deux grands types d'erreurs dans les statistiques produites à l'aide de données administratives : les erreurs de mesure et les erreurs de représentation. Le premier type se rapporte aux erreurs de mesure des caractéristiques (par exemple, l'âge, le sexe, etc.), tandis que le second a trait aux erreurs de représentation des unités ou des objets démographiques (par exemple, les individus ou les ménages dans un recensement)⁶. Zhang

⁵ Par exemple, les ajustements effectués sur la base de l'enquête postdénombrement réalisée après le recensement américain de 1990 ont fait l'objet de procédures judiciaires et ont été rejetés par la Cour suprême des États-Unis ; les ajustements de 2000 ont également fait l'objet de litiges (United States Census Bureau 2009).

⁶ D'après Zhang (2012), pour ce qui est des données d'entrée, les erreurs de mesure sont liées aux différences entre les caractéristiques fournies et les caractéristiques cibles (par exemple, le genre, le sexe, l'âge, l'origine ethnique, la profession, etc.) et comprennent plusieurs types d'erreurs au sein

fait également la distinction entre la qualité des sources uniques transmises par les fournisseurs de données et la qualité des sources transformées et/ou intégrées, après traitement par l'organisme national de statistique. Cette approche est reprise dans les Directives, qui évaluent la qualité des sources administratives uniques et des sources intégrées, en mettant l'accent sur la détection des erreurs de mesure et de représentation.

16. En outre, le système de l'erreur totale dans les enquêtes a également été adapté pour permettre d'évaluer la qualité des données administratives. Contrairement à l'ESM, l'EST recouvre un plus large éventail d'erreurs, notamment les erreurs de validité, de cadre ou de couverture, de non-réponse, de mesure, de traitement et de modèle. Ainsi, les modèles fondés sur l'erreur totale dans les enquêtes ont cherché à déterminer comment diverses erreurs s'accumulent tout au long de la conception et de la mise en œuvre de la méthode statistique, pour aboutir à l'erreur finale d'une estimation donnée. Cette approche a été adaptée pour rendre compte de la qualité des statistiques qui intègrent des données administratives (par exemple, Reid, Zabala et Holmberg 2017, Rogers et Blackwell 2020). Cependant, la qualité des statistiques ne peut être réduite à la seule évaluation des erreurs. Lorsqu'on envisage d'intégrer des données provenant d'une source administrative dans le plan de recensement, les conséquences de cette intégration sur la qualité doivent être évaluées en fonction de la mesure dans laquelle cette opération ajoute une erreur ou une incertitude aux résultats, par rapport aux avantages qu'elle présente, par exemple la réduction de la charge de travail pour les répondants, l'augmentation de la rapidité d'exécution et la réduction des coûts. Par conséquent, les présentes Directives définissent d'autres aspects susceptibles d'influer sur la qualité globale des résultats du recensement, notamment l'environnement institutionnel et la nécessité d'équilibrer les critères de qualité afin de répondre aux besoins des utilisateurs.

17. Le respect des Directives contribuera à garantir que les estimations du recensement soient fondées sur les sources et les méthodes les plus appropriées et ne prêtent pas à confusion. En même temps, il faut aussi tenir compte de la façon dont les sources administratives doivent être utilisées dans la conception du recensement. Étant donné la diversité des utilisations possibles, ce cadre doit être utilisé avec souplesse et être adapté au niveau de qualité que requièrent les différentes utilisations des données administratives par l'organisme national de statistique et aux différentes exigences des utilisateurs des statistiques de recensement, notamment le grand public, les organisations et les administrations locales et nationales. L'évaluation de la qualité repose donc inévitablement sur un jugement professionnel qualifié tout au long du processus de production statistique, de la collecte à la publication, afin de répondre aux besoins des utilisateurs.

B. Mesurer la qualité

18. La qualité des estimations de recensement produites à partir de sources administratives est particulièrement difficile à évaluer et/ou à mesurer en raison de la complexité et du caractère multidimensionnel des données utilisées. Comme indiqué ci-dessus, de nombreux facteurs influant sur la qualité ne sont pas mesurables quantitativement. En outre, ce qui constitue l'adéquation avec le but poursuivi et la haute qualité des statistiques variera nécessairement d'un utilisateur à l'autre ; par exemple, certains utilisateurs peuvent privilégier le degré d'actualité par rapport à la précision. Il est donc important d'évaluer ou de mesurer la qualité des données administratives dans les dimensions clés qui intéresseront les producteurs et les utilisateurs de statistiques. Par conséquent, il convient de préciser ce que l'on entend par évaluation et mesure.

des variables, notamment les erreurs de pertinence (défaut de concordance des définitions), de mappage (erreurs de reclassement des mesures dues à une mauvaise équivalence entre les classifications fournies et les classifications cibles, qui peuvent donc nécessiter des ajustements, par exemple par imputation) et de comparabilité (erreurs entre les mesures reclassées et ajustées). Les erreurs de représentation sont liées à la différence entre les unités fournies et les unités cibles. Elles comprennent les erreurs liées à la sur-couverture et à la sous-couverture (manque de concordance avec la population cible), à l'identification (erreurs de classification d'une unité dues à des incohérences entre plusieurs sources) et aux erreurs d'unité (erreurs dans la création d'unités statistiques d'intérêt lorsque celles-ci n'existent dans aucune source de données disponible).

19. Les présentes Directives font la distinction entre *l'évaluation de la qualité*, c'est-à-dire une évaluation qualitative, et *la mesure de la qualité*, c'est-à-dire le fait d'attacher une mesure quantitative à cette évaluation de la qualité. Lorsqu'il n'est pas possible de produire des indicateurs pour la mesure quantitative, ou que de tels indicateurs n'ont pas encore été mis au point, les Directives recommandent de procéder à une évaluation qualitative. En outre, plusieurs principes supplémentaires guident la production de statistiques officielles (UNECE 1992) et sont applicables à l'ensemble du processus statistique et à l'environnement plus large des organismes nationaux de statistique (par exemple, l'engagement en faveur de la qualité, de l'indépendance, de la protection des données, du secret statistique, etc.). Ces thèmes sont importants pour tous les processus statistiques et n'entrent pas entièrement dans le cadre des présentes Directives. Il faut toutefois reconnaître qu'un recensement qui utilise des sources administratives repose sur des données produites en dehors du système statistique, dans une organisation différente sur laquelle l'organisme national de statistique n'a généralement aucun contrôle⁷. Pour cette raison, les conséquences de l'utilisation de ces sources extérieures sur ces principes doivent être examinées avec soin.

C. Les étapes d'évaluation de la qualité

20. Pour que les Directives soient faciles à suivre, l'évaluation de la qualité des sources administratives est considérée au cours de quatre grandes étapes du cycle de vie du recensement. Ces étapes sont applicables quel que soit le type de recensement. Bien que la conception statistique ne soit jamais entièrement linéaire, le fait de réfléchir ainsi à la manière de procéder à l'évaluation de la qualité devrait permettre aux producteurs de statistiques de cerner rapidement les principales considérations relatives à la qualité qui correspondent le mieux à leur propre situation. Les étapes sont les suivantes :

- *Étape relative aux sources* : Une évaluation de la qualité des sources administratives nouvelles ou renouvelées à utiliser dans le recensement, basée sur les métadonnées. Cette étape ne nécessite pas que les organismes nationaux de statistique soient en possession des données réelles, mais elle est essentielle pour passer aux étapes suivantes ;
- *Étape relative aux données* : L'évaluation de la qualité des données administratives brutes fournies aux organismes nationaux de statistique par les autorités administratives. Pour ce faire, les organismes nationaux de statistique devront valider les données fournies par rapport aux enseignements tirés de l'étape relative aux sources. Outre la validation de base, cette étape comprend tout traitement nécessaire pour établir la qualité des données fournies par rapport à ce qui était attendu, ainsi que des comparaisons avec d'autres sources ;
- *Étape relative aux processus* : Les processus souvent appliqués aux sources de données administratives, dans le contexte des recensements, pour transformer les données en vue de leur utilisation dans le recensement et/ou pour en améliorer la qualité. Les processus considérés comprennent le couplage des données, la construction de registres statistiques et de la méthode des signes de vie, le dénombrement à l'aide de données administratives, les méthodes de comparaison de la qualité des variables d'une source à l'autre, la vérification et l'imputation ;
- *Étape relative aux résultats* : L'évaluation globale de la qualité des résultats de recensement obtenus à partir de données administratives. Bien que, d'un point de vue conceptuel, cette étape ne soit pas très différente de l'évaluation des résultats d'un recensement traditionnel, les présentes Directives tentent de cerner les différences éventuelles.

⁷ Dans certains cas, l'organisme national de statistique exerce un certain contrôle sur le registre. En Suisse, par exemple, le registre fédéral des bâtiments et des logements ou le registre des entreprises sont des sections de l'Office fédéral de la statistique. Par conséquent, il pourrait être possible, dans une perspective à long terme, d'intégrer certains registres appropriés au sein des organismes nationaux de statistique.

D. Les critères de qualité

21. Comme indiqué précédemment, la qualité des statistiques et des données administratives englobe de multiples dimensions qui ne se réduisent pas à des erreurs de représentation ou de mesure : par exemple, des statistiques précises mais obsolètes sont d'une utilité limitée. Les critères de qualité définis par le SSE comprennent 1) la pertinence, 2) la précision et la fiabilité, 3) le degré d'actualité et de ponctualité, 4) l'accessibilité et la clarté, et 5) la cohérence et la comparabilité⁸. Toutefois, pour l'évaluation des données administratives, ces critères habituels de qualité ne sont pas toujours applicables (Daas et al 2008, p.2). En revanche, ils permettent de prendre en compte tous les aspects pertinents de la qualité des données administratives. Les tableaux suivants présentent les critères d'évaluation des sources administratives utilisés dans les présentes Directives, pour chacune des étapes susmentionnées.

Tableau 1
Critères de qualité au cours de l'étape relative aux sources

<i>Critère de qualité</i>	<i>Définition</i>
Pertinence et précision	Mesure dans laquelle la source de données administratives répond aux besoins du recensement. Prise en compte du chevauchement entre la population cible du recensement, les concepts et les définitions (pertinence). Mesure dans laquelle les données décrivent correctement le phénomène qu'elles sont censées mesurer (précision).
Degré d'actualité	Laps de temps entre la fin de la période de référence à laquelle les informations se rapportent et la date à laquelle les informations sont mises à la disposition de l'organisme national de statistique
Cohérence et comparabilité	Mesure dans laquelle la source administrative peut être combinée avec succès avec d'autres sources utilisées dans le recensement, y compris la capacité de couplage.
Accessibilité	Facilité avec laquelle l'organisme national de statistique peut obtenir les données administratives, y compris les conséquences de toute restriction, l'acceptabilité publique de l'utilisation, la facilité de transfert et de réception des données et la disponibilité des métadonnées.
Environnement institutionnel	Facteurs organisationnels influant sur la capacité du fournisseur de données à fournir des données de la qualité attendue, y compris la solidité de la relation, l'expérience antérieure, l'existence d'accords formels, les risques associés au statut du fournisseur et les normes de qualité du fournisseur.

⁸ D'autres critères sont utilisés par divers organismes nationaux de statistique (par exemple, Statistique Canada 2017, Australian Bureau of Statistics 2009). Dans l'ensemble, ces autres cadres couvrent approximativement le même contenu, bien qu'ils utilisent une terminologie ou des classifications différentes.

Tableau 2
Critères de qualité au cours de l'étape relative aux données

<i>Critère de qualité</i>	<i>Définition</i>
Validation et harmonisation	Les fichiers de données fournis à l'organisme national de statistique sont dans un format lisible. D'autres dispositifs de validation et d'harmonisation des données sont en place lors du transfert des données à l'organisme national de statistique, qui permettent de confirmer que les variables, unités ou périodes de référence prévues ont été fournies et de garantir que le traitement des données par l'organisme national de statistique est cohérent dans tous les cas d'utilisation du recensement.
Précision et fiabilité	La précision, l'exhaustivité (pour les variables et la couverture de la population) et la cohérence des données fournies correspondent aux exigences du cas spécifique de recensement pour lequel ces données seront utilisées. Les comparaisons avec d'autres sources révèlent des niveaux acceptables d'erreurs de mesure ou de représentation.
Degré d'actualité et de ponctualité	Le degré d'actualité et de ponctualité des données fournies correspond aux exigences du cas spécifique de recensement pour lequel ces données seront utilisées.
Capacité de couplage	Des variables de couplage adéquates sont disponibles (c'est-à-dire soit des identifiants uniques communs, soit une combinaison de variables permettant l'identification) et leur qualité est suffisante pour permettre le couplage de données.

Tableau 3
Critères de qualité au cours de l'étape relative aux processus

<i>Critère de qualité</i>	<i>Définition</i>
Précision du couplage des informations	Lorsque des sources multiples sont liées (entre elles ou aux réponses au recensement), le couplage est précis et non biaisé, ce qui améliore la qualité globale de la méthode de recensement et/ou de l'ensemble de données.
Champ et cohérence des registres statistiques et des dénombrements administratifs	Lorsque des registres de (sous-)population de recensement sont construits, ou lorsque des données administratives sont utilisées pour compléter la collecte de données de recensement, ces registres ou données couvrent de manière adéquate la population ou les variables cibles, améliorant ainsi la qualité globale de la méthode de recensement et/ou de l'ensemble de données de recensement.
Précision de la résolution des conflits	Lorsque différentes sources sont liées et que les mêmes attributs sont disponibles dans ces sources, les méthodes permettant de trancher entre les sources améliorent la qualité globale de la méthode de recensement et/ou de l'ensemble de données de recensement.
Précision de la vérification et de l'imputation	Lorsque les variables ou les unités de recensement sont dérivées ou construites à l'aide de techniques d'imputation ou de modélisation, cette dérivation est précise et non biaisée, améliorant ainsi la qualité globale de la méthode de recensement et/ou de l'ensemble de données de recensement.

Tableau 4
Critères de qualité au cours de l'étape relative aux résultats

<i>Critère de qualité</i>	<i>Définition</i>
Pertinence	Mesure dans laquelle les résultats statistiques répondent aux besoins actuels et potentiels des utilisateurs.
Précision et fiabilité	Proximité entre un résultat estimé et la valeur réelle inconnue – et fiabilité de ces éléments dans le temps et dans l'espace.
Degré d'actualité et de ponctualité	Laps de temps entre la publication et la période à laquelle les données se rapportent et décalage entre la date de publication effective et la date prévue.
Accessibilité et clarté	L'accessibilité et la clarté font référence aux mesures prises pour aider l'utilisateur à trouver et à comprendre les données qui l'intéressent.
Cohérence et comparabilité	Mesure dans laquelle les données peuvent être comparées dans le temps et entre différents domaines. Degré de similitude entre des données provenant de sources ou de méthodes différentes, mais se rapportant à un même phénomène.

Source : Eurostat 2013 et 2018.

III. Étape relative aux sources

22. Ce chapitre présente les principaux critères de qualité, le processus d'évaluation, ainsi que les outils et indicateurs associés pour évaluer la qualité des sources de données administratives à utiliser dans le cadre de la production du recensement, tant lors de la première obtention des sources que lorsque celles-ci sont régulièrement remises à la disposition de l'organisme national de statistique. Normalement, aucune donnée n'est accessible au cours de cette étape. Toutefois, la recherche d'informations sur les sources de données administratives commence, très probablement par des communications et des réunions exploratoires entre l'organisme national de statistique et le fournisseur de la source de données administratives.

23. L'évaluation réalisée au cours de cette étape doit conduire à une recommandation sur l'opportunité de poursuivre la démarche d'obtention (ou de prolonger la fourniture de la source de données). Si la décision est prise d'aller dans ce sens, les données administratives fournies feront l'objet d'une évaluation plus détaillée lors de l'étape relative aux données.

24. Il est nécessaire d'évaluer la qualité des sources à la fois lors de la première obtention d'une source de données administratives et chaque fois que celle-ci est à nouveau mise à la disposition de l'organisme national de statistique. En effet, les caractéristiques de tout ensemble de données qui a été fourni précédemment peuvent avoir subi des changements en ce qui concerne les concepts, la classification, les méthodes de collecte, etc.

A. Critères de qualité des sources

25. Les critères de qualité des données à considérer au cours de cette étape d'évaluation de la qualité sont la pertinence, le degré d'actualité, la cohérence et la comparabilité, l'accessibilité et l'interprétabilité, ainsi que l'environnement institutionnel. Ces critères sont décrits ci-après et les processus, outils et indicateurs utilisés pour l'évaluation sont présentés dans les sections suivantes. Il convient de noter que l'incapacité à atteindre une qualité minimale acceptable au regard de l'un des critères ne peut être compensée par le respect des autres critères.

1. Pertinence et précision

26. La pertinence est la mesure dans laquelle une source de données administratives répond aux besoins de l'organisme national de statistique par rapport à l'utilisation prévue. Pour être jugée pertinente, cette source de données doit, par exemple, répondre aux raisons pour lesquelles elle a été obtenue. Il peut s'agir d'une réduction des coûts ou de la charge de travail des répondants, d'une amélioration de la qualité des résultats du recensement ou de la fourniture de résultats de recensement améliorés ou nouveaux. Pour cela, la source administrative doit être représentative de la population qui présente un intérêt pour le recensement (la population cible) et les mesures effectuées auprès de cette population doivent correspondre aux besoins du recensement. Pour évaluer la pertinence, il est essentiel de comprendre le contexte dans lequel les données administratives ont été recueillies.

27. Dans le cadre de l'évaluation de la pertinence, la précision des données administratives est également prise en considération. La précision fait référence à la mesure dans laquelle les données décrivent correctement le phénomène qu'elles sont censées mesurer. Il est important de comprendre comment la collecte, le traitement et l'assurance qualité effectués par l'organisation administrative peuvent influencer sur la précision des données obtenues et sur leur utilité.

2. Degré d'actualité

28. Le degré d'actualité fait référence à la période entre la date à laquelle les informations se rapportent et la date à laquelle ces informations sont mises à la disposition de l'organisme national de statistique. Le degré d'actualité des informations influera sur leur pertinence.

3. Cohérence et comparabilité

29. La cohérence est la mesure dans laquelle les données administratives peuvent être combinées avec succès, au fil du temps, avec les données provenant d'autres sources utilisées par l'organisme national de statistique, c'est-à-dire les données de recensement, dans un cadre analytique général. L'utilisation de concepts, de classifications et de populations cibles normalisés favorise la cohérence dans les recensements et entre eux. Par conséquent, une compréhension claire des définitions opérationnelles utilisées par le fournisseur administratif, l'objectif de la collecte de données et l'incidence des changements apportés à une source administrative sur la comparabilité dans le temps et entre différents domaines jouent un rôle dans l'évaluation de la cohérence.

30. Il est souvent nécessaire de relier une source administrative au niveau de l'unité statistique de recensement pour intégrer les données dans la conception du recensement. La comparabilité des identifiants entre les différentes sources de données à relier est donc un élément à prendre en compte dans le cadre de la cohérence.

4. Accessibilité et interprétabilité

31. L'accessibilité désigne la facilité avec laquelle l'organisme national de statistique peut obtenir (et comprendre) les éléments de données administratives pertinents dans leur intégralité. Cela inclut la compréhension de toutes les restrictions (qu'elles soient légales ou imposées par le fournisseur), l'acceptation par le public, la facilité de transfert et de réception des données (forme et support adaptés au transfert des données et coûts), ainsi que la disponibilité et la clarté de la documentation et des métadonnées. Il est essentiel que l'utilisation de la source de données administratives repose sur un cadre juridique qui donne à l'organisme national de statistique le droit non équivoque d'accéder aux données et aux métadonnées et de les utiliser à des fins statistiques.

5. Environnement institutionnel

32. L'environnement institutionnel fait référence aux facteurs organisationnels ou institutionnels qui peuvent avoir une incidence sur la capacité du fournisseur de données à fournir des données de la qualité attendue et selon le calendrier convenu (ponctualité). Il s'agit notamment de la solidité de la relation avec le fournisseur de données, y compris l'efficacité des canaux de communication et la réactivité du fournisseur aux demandes de l'organisme national de statistique. Cela inclut l'existence (ou la possibilité) d'accords

formels et les risques associés à la nature et à la complexité de l'organisation du fournisseur. Cela comprend également les normes et procédures de qualité adoptées par la ou les organisations administratives.

B. Outils et indicateurs

33. La qualité d'une source administrative doit être évaluée en fonction des critères de qualité décrits dans la section ci-dessus. On trouvera ci-après des conseils sur le processus d'évaluation, ainsi que des outils et des indicateurs permettant d'évaluer une source administrative en vue de son utilisation dans le recensement.

1. Pertinence et précision

34. Il est important, pour évaluer la pertinence et la précision, de comprendre les différences entre la population administrative et la population à recenser, ainsi qu'entre les mesures ou les variables de la source administrative et les caractéristiques à prendre en compte dans le recensement. L'erreur découlant de ces différences est appelée respectivement erreur de représentation et erreur de mesure (Zhang, 2012). Au stade de l'évaluation de la source, il est possible d'acquérir une certaine compréhension de ces erreurs et de leur incidence sur la pertinence en se fondant sur les métadonnées relatives à l'objectif et aux méthodes de collecte des données du fournisseur. Les conséquences des erreurs de représentation et de mesure sur la précision et la fiabilité sont également prises en compte lors des étapes relatives aux données et aux processus.

a) La population cible du recensement (représentation)

35. Pour évaluer la pertinence, l'organisme national de statistique doit déterminer si l'ensemble des objets d'une source de données administratives correspond aux unités de population qui présentent un intérêt pour le recensement (la population cible). Un objet est l'élément de base de la population sur lequel des informations sont collectées, il peut s'agir d'une personne, d'un ménage, d'un logement, d'un événement ou de transactions, etc. Les indicateurs qui suivent sont proposés pour établir la pertinence, du point de vue de la représentation. Chaque indicateur est assorti d'une série de questions destinées à guider l'évaluation :

- Correspondance (des objets) avec les unités cibles du recensement :
 - i) Dans quelle mesure les objets administratifs sont-ils comparables aux unités cibles du recensement ?
 - ii) Quelles définitions, méthodes et processus sont utilisés pour distinguer un objet et l'inclure dans la source ?
 - iii) Existe-t-il des lois ou des règlements qui définissent les objets ?
 - iv) Des contrôles sont-ils effectués par le détenteur des données pour s'assurer de la validité des définitions ?
 - v) En cas de défaut de correspondance avec les unités de recensement, une transformation est-elle possible qui pourrait répondre aux besoins du recensement ?
- Couverture (de l'ensemble des objets) par rapport à la population cible du recensement :
 - i) La couverture des objets répond-elle aux besoins du recensement ?
 - ii) Existe-t-il des cas de sous-couverture (objets ne figurant pas dans la source, mais faisant partie de la population cible du recensement) et de surcouverture (objets figurant dans la source, mais ne faisant pas partie de la population cible du recensement) qui auraient une incidence sur l'utilité de la source ?
 - iii) Existe-t-il des différences entre les zones géographiques dues à des différences dans les pratiques du détenteur de données ou à la législation qui doivent être prises en compte ?

iv) Existe-t-il des règles ou des exigences législatives ou réglementaires, y compris des sanctions pour non-conformité, qui peuvent avoir une incidence sur l'inclusion d'objets dans la source ou sur leur exclusion ?

v) Quels sont les méthodes et processus adoptés par le détenteur de données pour inclure de nouveaux objets qui répondent aux critères d'inclusion ou aux définitions requis (par exemple, des procédures d'enregistrement) et pour supprimer les objets qui ne correspondent plus à la population cible en ce qui concerne la source administrative (par exemple, des procédures de radiation) ?

vi) Dans le cas d'erreurs de couverture, existe-t-il d'autres sources de données qui pourraient être utilisées en combinaison avec la source pour remédier, par exemple, à une sous-couverture ou à une surcouverture dans la source ?

b) Les variables ou les caractéristiques du recensement (mesure)

36. Pour évaluer la pertinence, l'organisme national de statistique doit également déterminer si les informations recueillies à partir des objets sur une source de données administratives répondent aux besoins du recensement, par rapport aux concepts cibles (par exemple, l'appartenance ethnique, la situation professionnelle, la taille du ménage, le statut d'occupation, etc.) Les indicateurs qui suivent sont proposés pour établir la pertinence, en ce qui concerne la mesure :

- Disponibilité des variables ou des caractéristiques cibles :
 - i) La source administrative comprend-elle les variables nécessaires par rapport à l'utilisation prévue pour le recensement ?
 - ii) Les variables ou les caractéristiques couvrent-elles globalement la période de référence du recensement ?
- Correspondance des concepts, des définitions et de la classification des variables avec les besoins du recensement :
 - i) Les concepts, définitions et classifications administratifs sont-ils comparables aux besoins du recensement ?
 - ii) Existe-t-il une différence entre les concepts cibles idéaux du détenteur de données et ce qui est effectivement obtenu par la mesure opérationnelle cible de ces concepts utilisée dans la collecte ?
 - iii) En cas de défaut de correspondance avec les concepts, définitions et classifications du recensement, une transformation est-elle possible pour répondre aux besoins du recensement ?
- Erreur de correspondance ou de mesure par rapport à la période de référence du recensement :
 - i) Quelle est la fréquence de collecte d'une variable ou d'une caractéristique ?
 - ii) Y a-t-il des délais connus entre la survenue d'un événement ou d'un phénomène et son intégration dans la source administrative (par exemple, les parents peuvent ne pas avoir à enregistrer une naissance avant plusieurs semaines dans le registre des naissances d'un pays) ?
 - iii) Des horodatages sont-ils enregistrés dans la source de données pour indiquer à quelle période se rapporte une donnée ?
 - iv) Existe-t-il des mesures de nature à inciter une personne concernée à mettre à jour ses informations dans la source administrative au fur et à mesure que sa situation ou ses informations changent ou à l'en dissuader (par exemple, des avantages ou des pénalités pour ne pas le faire ou pour le faire) ?
- Qualité de la collecte et risques d'erreur de mesure par rapport aux concepts du recensement :
 - i) Quel est le but poursuivi par le détenteur de données pour collecter celles-ci et comment cela peut-il influencer sur la qualité des données ?

- ii) Existe-t-il des obligations légales, des objectifs ou des incitations (ou une absence d'incitations) qui pourraient influencer sur la qualité de la collecte des données ?
 - iii) Le processus de collecte suivi par le détenteur de données soulève-t-il des inquiétudes quant à la qualité des variables, y compris quant au risque de distorsion ? Il peut s'agir notamment de l'enregistrement des données par un intermédiaire et donc de l'absence de déclaration directe par la personne concernée (ce qui accroît le risque de fausse déclaration) ;
 - iv) Quelles sont les procédures en place pour valider et vérifier les données lors de leur saisie par le détenteur de données ?
 - v) Existe-t-il des mesures de nature à inciter les personnes concernées à fournir des informations détaillées et précises au détenteur de données ou à les en dissuader ?
- Qualité du traitement des données et risques d'erreur de traitement par le détenteur de données :
 - i) Le traitement effectué par le détenteur de données permet-il de penser que la qualité des données obtenues répondra aux besoins du recensement ?
 - ii) Quels contrôles sont effectués par le détenteur de données pour garantir la qualité ?
 - iii) Les données sont-elles vérifiées ou imputées ? Si c'est le cas, quand et comment, et existe-t-il dans la source de données un indicateur permettant de savoir si une vérification ou une imputation a eu lieu ?
 - iv) Le détenteur de données est-il soumis à des règles, des réglementations ou des incitations susceptibles d'avoir une incidence sur la manière dont les données sont traitées ?

37. Au cours de l'étape relative aux sources, l'évaluation par rapport aux indicateurs repose généralement sur une évaluation qualitative (indiquant par exemple, sur la base des réponses aux questions posées, si le besoin est entièrement satisfait, l'est partiellement ou ne l'est pas pour chaque indicateur, avec une explication des raisons à cela). Une évaluation quantitative de l'erreur de représentation ou de mesure est effectuée au cours de l'étape relative aux données (sur la base de l'analyse des données) selon le critère de la précision et de la fiabilité.

38. L'évaluation par rapport aux indicateurs doit permettre de prendre une décision (souvent fondée sur l'expérience et le jugement d'experts) sur l'utilisation (ou la poursuite de l'utilisation) d'une source dans le recensement. Cette décision doit tenir compte de la possibilité que la source de données réponde aux besoins du recensement (réduction des coûts et de la charge de travail des répondants, amélioration et renforcement des résultats du recensement, par exemple), par rapport à l'ensemble des coûts ou des risques (mentionnés ci-après à propos des critères relatifs à l'environnement institutionnel et à l'accessibilité).

39. Il existe divers cadres de qualité décrits dans la littérature qui fournissent des indicateurs similaires à ceux qui sont présentés dans le présent chapitre en fonction des différents critères de qualité, ainsi que des ensembles de questions et des systèmes de notation qui permettent d'éclairer l'évaluation (par exemple, Daas *et al.* 2009 ; SN-MAID, 2014 ; Iwig *et al.* 2013 ; Guide d'évaluation des données administratives de Statistique Canada (Lavigne & Nadeau 2014) ; Statistics Austria's Quality Assessment of Administrative Data, Documentation of Methods Framework (Statistics Austria 2019)).

2. Degré d'actualité

40. Une source administrative peut couvrir la période correspondant au recensement, mais pour être utile, elle devra également être disponible à temps, en fonction du calendrier du recensement. L'indicateur qui suit est proposé pour évaluer le degré d'actualité :

- Degré d'actualité et fréquence de l'offre par rapport aux besoins du recensement :
 - i) Quel est le délai entre la survenue d'un événement ou d'un phénomène et sa saisie dans la source administrative ?

ii) Quel est le délai entre la fin de la période de référence concernant les données administratives et la date à laquelle ces données peuvent être mises à la disposition de l'organisme national de statistique ?

iii) À quelle fréquence les données peuvent-elles être fournies à l'organisme national de statistique, compte tenu des besoins du recensement ?

iv) Existe-t-il des exigences, quant à la méthode de fourniture, aux formats requis et aux structures de données utilisées par l'organisme national de statistique qui pourraient avoir une incidence sur le calendrier du fournisseur de données ?

v) L'organisme national de statistique peut-il structurer et traiter les données à temps pour répondre aux besoins du recensement, à partir du moment où les données sont disponibles ?

41. Dans les cas où il est peu probable que les données soient disponibles à temps, l'organisme national de statistique peut souhaiter déterminer si une version provisoire de l'ensemble de données peut être mise à sa disposition avant la date prévue. Dans ce cas, l'ensemble de données peut être incomplet et sujet à des niveaux d'erreur plus élevés. Il peut donc s'avérer nécessaire de trouver un compromis entre le degré d'actualité des données et leur précision.

42. Comme indiqué ci-après à propos du critère relatif à l'environnement institutionnel, il est important d'inclure les dates de livraison par rapport aux périodes de référence des données dans les accords formels conclus avec le fournisseur de données. Bien qu'elles puissent être disponibles à temps pour répondre aux exigences du propriétaire des données, les données ne seront pas nécessairement fournies à temps à l'organisme national de statistique, alors que ce dernier est formellement responsable de la fourniture en temps voulu des données de recensement.

3. Cohérence et comparabilité

43. Il est important d'évaluer dans quelle mesure une source administrative peut être combinée avec succès avec d'autres sources de données pour être utilisée dans le recensement. Les informations recueillies selon les indicateurs fournis pour évaluer la pertinence peuvent être utilisées pour évaluer la cohérence. Il s'agit notamment d'informations sur les différences entre les concepts, les définitions, les classifications et les méthodes sous-jacentes de la source de données administratives et des autres sources de données à utiliser conjointement dans le cadre du recensement.

44. Pour un recensement entièrement basé sur les registres, il est important d'analyser les caractéristiques du recensement et la source de données administratives ; il s'agit de déterminer et de vérifier dans quelle mesure les informations contenues dans la source de données administratives facilitent la dérivation des caractéristiques pertinentes du recensement. En particulier, l'organisme national de statistique doit déterminer si les données enregistrées dans les registres sont conformes ou non à la définition des caractéristiques du recensement. En cas de conformité partielle ou nulle, l'organisme national de statistique doit examiner les causes de la non-conformité entre les caractéristiques du recensement et les informations disponibles dans la source de données administratives.

a) Comparabilité

45. Les données administratives sont sujettes à des variations et à des différences dans le temps et d'un domaine à l'autre dues à des changements dans la législation, la réglementation et les procédures, qui peuvent influencer sur les concepts, les définitions, les classifications et la couverture d'une source. Plus généralement, ces changements peuvent avoir une incidence sur tous les indicateurs de représentation et de mesure, comme indiqué à propos du critère de pertinence. Ceci est particulièrement important pour le recensement, où la stabilité dans le temps peut être une préoccupation majeure. L'indicateur qui suit est proposé pour évaluer la comparabilité :

- Comparabilité dans le temps et entre différents domaines :

Y a-t-il des changements dans le temps ou des différences entre les domaines (par exemple, les zones géographiques) qui influent sur :

- i) La définition d'un objet et la couverture des objets figurant dans la source administrative utile au recensement ?
- ii) Les concepts, définitions et classifications associés aux variables figurant dans la source administrative utile au recensement ?
- iii) Les procédures de collecte, de traitement et d'assurance qualité des données qui pourraient avoir une incidence sur la qualité de la source utilisée pour le recensement ?

b) Capacité de couplage

46. Un aspect qui relève de la cohérence et de la comparabilité est la facilité avec laquelle une source de données administratives peut être reliée à d'autres ensembles de données utiles au recensement. Les indicateurs suivants sont proposés pour évaluer la capacité de couplage d'une source :

- Présence d'une clef unique permettant le couplage :
 - i) La source comprend-elle un identifiant unique qui est commun avec les clefs uniques requises pour le couplage du recensement ?
 - ii) L'identifiant est-il disponible pour tous les objets pertinents de la source ?
- Présence d'une combinaison unique de variables permettant le couplage :
 - i) La source comprend-elle une combinaison unique de variables (par exemple, le nom, l'âge et l'adresse) qui pourrait être utilisée pour le couplage du recensement ?
 - ii) La combinaison unique de variables est-elle présente pour chaque objet de la source ?

47. La qualité des variables de couplage est également évaluée lors de l'étape relative aux données et la qualité du processus de couplage est examinée dans le cadre de l'étape relative aux processus.

4. Accessibilité

48. Les indicateurs qui suivent sont proposés pour l'évaluation en fonction du critère d'accessibilité :

- Restrictions portant sur l'accès aux données et leur utilisation ;
- Acceptabilité par le public ;
- Facilité de transfert et de réception des données ;
- Interprétabilité de la source –métadonnées claires et détaillées.

49. Les sections ci-après contiennent des informations détaillées concernant l'évaluation en fonction de chacun des indicateurs.

a) Restrictions portant sur l'accès aux données et leur utilisation

50. Il est important de mettre en évidence toutes les restrictions qui peuvent avoir une incidence sur la capacité de l'organisme national de statistique d'accéder à une source administrative et d'utiliser cette source. Par exemple, les restrictions liées à la protection des données inscrites dans des actes juridiques peuvent imposer certaines limites à l'acquisition et au traitement des données (en particulier lorsque celles-ci sont protégées par des mesures ou des lois de sécurité supplémentaires au niveau de l'État). Ces actes juridiques peuvent être spécifiques à des sources de données particulières ou être plus génériques, permettant à l'organisme national de statistique d'accéder à ces sources de données selon les besoins, sous réserve de l'accord du propriétaire des données. Celui-ci peut également imposer d'autres restrictions à la fourniture des données et à leur utilisation autorisée. Ces restrictions peuvent inclure :

- La suppression de registres ou de variables ;

- Des traitements à des fins de divulgation (avant la livraison), comme le cryptage des identifiants ou la déstructuration, le groupement ou le plafonnement des données fournies ;
- Des restrictions portant sur la manière dont les données peuvent être utilisées ;
- Des restrictions relatives à la rétention des données et des règles de suppression ou d'élimination ;
- Des règles concernant les méthodes de divulgation qui doivent être appliquées par l'organisme national de statistique, qui ont une incidence sur les résultats du recensement.

51. L'organisme national de statistique doit stipuler et décrire toutes les restrictions qui s'appliquent, ce qui peut donner lieu à une évaluation des conséquences (et des risques) liés aux restrictions relatives à l'utilisation d'une source administrative dans le recensement. Dans le cadre de cette évaluation, il doit également déterminer s'il a la capacité de respecter les restrictions. Il peut notamment devoir adopter des garanties techniques et procédurales. Ces garanties feront généralement partie d'un protocole d'accord ou d'un accord sur la sécurité des données conclu avec le propriétaire ou le fournisseur des données. En particulier, le protocole d'accord peut préciser comment les informations permettant d'identifier les personnes seront protégées.

b) Acceptabilité par le public

52. La possibilité pour un organisme national de statistique d'accéder à une source de données pour l'utiliser dans le cadre d'un recensement peut également dépendre de l'acceptation du public. Celui-ci doit comprendre et soutenir les nouvelles approches et utilisations de ses informations, ou du moins ne pas y être hostile. S'il s'oppose à l'utilisation d'une source de données administratives, il y a un risque pour la qualité. Par exemple, cela pourrait changer la façon dont le public interagit avec le recensement ou avec une source administrative utilisée dans le recensement. L'organisme national de statistique doit donc faire preuve de transparence quant à l'utilisation de sources administratives dans le recensement, en soulignant les avantages pour le public, tout en fournissant des garanties en matière de confidentialité et de sécurité.

53. Pour évaluer l'acceptabilité par le public, les outils ou processus suivants peuvent être utilisés :

- Une consultation ou une participation du public ;
- Une évaluation de l'impact sur la vie privée ;
- Une évaluation de l'éthique des données.

54. Un processus de consultation ou de participation du public peut être organisé par l'organisme national de statistique au sujet de l'utilisation des données administratives dans le recensement (ou pour d'autres recherches ou résultats statistiques). Ce processus peut prendre diverses formes, notamment des consultations formelles, des questionnaires (par le biais d'enquêtes ou par l'intermédiaire de l'organisme national de statistique, celui-ci invitant le public à donner son avis sur son site Web), des recherches qualitatives sur les attitudes du public, ou le recours à des groupes de citoyens. Les groupes de citoyens ont pour but de réunir des membres du public (pour représenter la population ou différents groupes d'intérêt) afin d'évaluer leurs points de vue et leurs opinions.

55. Une évaluation de l'impact sur la vie privée est un processus formel aboutissant à un document qui décrit le processus, les conclusions et les résultats et qui aide l'organisme national de statistique à examiner les effets d'un nouveau programme ou service (ou des politiques et plans proposés) sur la vie privée des personnes. En tant qu'outil de gestion des risques utilisé dans la phase de planification d'un programme ou d'un service, cette évaluation aide les organisations à mieux prendre en compte les conséquences d'une proposition donnée sur la vie privée. Les évaluations de l'impact sur la vie privée sont également utilisées pour s'assurer que les responsables du traitement des données peuvent remplir les obligations qui leur incombent en vertu du Règlement général sur la protection des données (dans le cadre du droit européen). Une telle évaluation peut être appliquée aux

différentes utilisations qu'un organisme national de statistique peut souhaiter faire d'une source de données lors de la conception du recensement.

56. Une évaluation de l'éthique des données a pour but d'établir si l'accès, l'utilisation et le partage des données publiques, à des fins de recherche et de statistiques, sont éthiques et servent les intérêts du public. Les organismes nationaux de statistique peuvent utiliser un outil d'auto-évaluation de l'éthique (UKSA 2020, par exemple), mais ils peuvent aussi faire appel à un organisme officiel pour obtenir des conseils d'experts ou une approbation, comme un comité consultatif sur l'éthique des données (par exemple, le Comité consultatif sur l'éthique des données du Statisticien en chef du Royaume-Uni⁹).

57. Les résultats de la consultation ou de la participation du public et des évaluations de l'impact sur la vie privée et de l'éthique des données peuvent aider un organisme national de statistique à évaluer l'acceptabilité par le public de l'utilisation de sources administratives pour le recensement (et pour d'autres statistiques de l'organisme national).

c) Facilité de transfert et de réception des données

58. Le fournisseur de données peut utiliser des modèles, formats et schémas de données, ainsi que des logiciels et du matériel de traitement des données, très différents de ceux que connaît l'organisme national de statistique. Ces différences peuvent porter sur la manière dont les données sont conservées et transmises (y compris les dispositions de sécurité applicables à la transmission). Les structures de données peuvent également être complexes et les fichiers extrêmement volumineux (en particulier pour les données de transaction). Pour évaluer s'il peut recevoir et intégrer les ensembles de données dans ses systèmes, il est important que l'organisme national de statistique comprenne ces différences et ces complexités. Ce processus peut également inclure des négociations avec le fournisseur sur la mise au point de processus et de systèmes visant à faciliter la transmission des ensembles de données sous un format répondant aux besoins de l'organisme national de statistique. Toutefois, un tel processus peut prendre du temps et être coûteux.

59. Plus généralement, le coût est un facteur clef à prendre en compte pour évaluer la facilité d'accès. Il peut s'agir de coûts imposés par le fournisseur de données ou de ceux que doit supporter l'organisme national de statistique pour se donner les moyens de recevoir un ensemble de données administratives (par exemple, si un logiciel ou du matériel nouveau doit être acheté). Il est important d'évaluer ces coûts par rapport à la valeur attendue d'une nouvelle source administrative.

60. Dans la pratique, les détails des modalités de transmission des données à l'organisme national de statistique, y compris les structures et le format des fichiers (fichiers plats, base de données relationnelle ; formats SAS, Excel ou texte, par exemple), les variables, la fréquence de fourniture et les dates de livraison, les normes relatives aux données et les coûts convenus, seraient inclus dans les accords de partage ou de fourniture de données entre l'organisme national de statistique et le fournisseur. Ces accords seraient signés par les responsables autorisés de chacune des organisations.

d) Interprétabilité de la source – métadonnées claires et détaillées

61. L'évaluation de l'interprétabilité porte sur l'existence et la disponibilité de métadonnées et de ressources documentaires complètes et claires sur la source administrative. Sans cela, il n'est pas possible de comprendre et d'évaluer la source administrative par rapport à l'utilisation prévue. Les métadonnées doivent inclure des détails sur :

- L'organisation administrative et l'objectif de la collecte ;
- Les concepts, définitions, classifications et protocoles utilisés ;

⁹ Pour de plus amples informations, voir <https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/national-statisticians-data-ethics-advisory-committee/#:~:text=The%20National%20Statistician%E2%80%99s%20Data%20Ethics%20Advisory%20Committee%20%28NSDEC%29,advise%20the%20National%20Statistician%20on%20the%20ethical%20>

- Les méthodes et procédures de collecte, de traitement, de validation et d'assurance qualité ;
- Les unités et variables de référence, y compris les dictionnaires de données, les structures de fichiers, les formats et les liens entre les données.

62. Ces informations sont importantes pour l'évaluation en fonction des autres critères de qualité décrits dans le présent chapitre. Il arrive souvent que des métadonnées claires et complètes n'existent pas pour tous les aspects d'une source administrative lors de la phase initiale d'exploration de cette source en vue de son utilisation par l'organisme national de statistique. Il est donc nécessaire de travailler avec le fournisseur de données pour construire les métadonnées pertinentes. Cette collaboration repose sur une bonne communication avec le fournisseur et sur la volonté de ce dernier de travailler avec l'organisme national de statistique (voir ci-après la section consacrée à l'environnement institutionnel). En fonction de la complexité d'une source administrative, un organisme national de statistique peut décider de détacher du personnel auprès de l'organisation administrative afin d'acquérir une connaissance approfondie de la source.

5. Environnement institutionnel

63. L'organisme national de statistique dépend entièrement du fournisseur de données pour que les données administratives soient collectées, traitées et livrées avec la qualité attendue et selon le calendrier convenu. Il est également tributaire de la qualité des informations que le fournisseur lui communique sur les données et sur toute modification prévue de ces dernières. Il est donc important d'évaluer la confiance dans la capacité du fournisseur de données de répondre à ces besoins. Les indicateurs qui suivent sont proposés pour l'évaluation de l'environnement institutionnel :

- La solidité de la relation avec le fournisseur ;
- Les expériences précédentes avec le fournisseur ;
- L'existence d'accords formels ;
- Le risque présenté par le statut du fournisseur ;
- Les normes de qualité du fournisseur.

64. *La solidité de la relation.* Des processus doivent avoir été mis en place pour gérer la relation avec le fournisseur de données et garantir un dialogue continu. Ces processus doivent inclure des mécanismes permettant :

- La communication des exigences de l'organisme national de statistique au fournisseur ;
- La communication en temps utile (par le fournisseur) de tout changement susceptible d'affecter la source de données (par exemple, des changements dans la base juridique des données, dans les concepts et les classifications et dans les processus et procédures de collecte, de gestion et de fourniture des données) ;
- De soulever toute question avec le fournisseur au sujet de la source de données ;
- De communiquer au fournisseur les résultats qui pourraient conduire à des améliorations de la source.

65. *Les expériences précédentes.* Il s'agit notamment de savoir dans quelle mesure un fournisseur a répondu aux demandes de l'organisme national de statistique, si des problèmes sont apparus lors de précédentes fournitures de données (livraison tardive, erreurs inattendues, par exemple), si le fournisseur a fourni par le passé des informations précises sur une source de données (cela pourrait avoir été établi par des vérifications effectuées à un stade ultérieur par l'organisme national de statistique).

66. *L'existence d'accords formels.* Il s'agit notamment de savoir si des accords écrits (juridiques ou autres) existent ou peuvent être élaborés, couvrant :

- Les rôles et responsabilités de l'organisme national de statistique et du fournisseur. Il peut être question de savoir si le service national de statistique a un rôle à jouer dans

l'approbation de tout changement apporté à une source administrative utilisée (ou devant être utilisée) dans le cadre du recensement ;

- La base juridique de la fourniture des données et les exigences en matière de sécurité et/ou de confidentialité ;
- La spécification des exigences, conformément à l'accord de partage ou de fourniture de données mentionné ci-dessus dans la section relative à la facilité de transfert et de réception des données.

67. *Le statut du fournisseur.* Le risque associé au statut d'un fournisseur doit être évalué par l'organisme national de statistique, qui doit examiner si le fournisseur est une organisation établie, stable et réputée. Il s'agit de déterminer si la fonction administrative exercée par le fournisseur repose sur une base légale ou réglementaire qui permette d'avoir confiance dans la durabilité et la qualité de la source. Les risques associés à la complexité de l'organisation ou des organisations du fournisseur impliquées dans la collecte, le traitement et la fourniture de la source doivent également être pris en compte. Par exemple, il peut y avoir plusieurs organismes ou organisations impliqués, chacun ayant une influence sur la qualité de la source de données finale.

68. *Les normes de qualité du fournisseur.* Il convient d'évaluer si le fournisseur peut répondre aux attentes de l'organisme national de statistique en matière de qualité. Cette évaluation doit prendre en compte les informations sur les principes, normes et directives adoptés par le fournisseur pour assurer la qualité, y compris la procédure en place couvrant la collecte, le traitement et la fourniture de données à l'organisme national de statistique. Les éléments indiquant comment le fournisseur vérifie le respect des normes sont précieux. Il peut s'agir d'audits internes ou externes réalisés par des autorités de réglementation ou des organismes professionnels. Le fournisseur peut également produire des rapports sur la qualité, qui doivent être examinés par l'organisme national de statistique. Une évaluation plus détaillée fondée sur les principaux aspects de la source administrative est incluse dans le critère de qualité relatif à la pertinence présenté plus haut.

69. Après avoir évalué le fournisseur de données en fonction des critères susmentionnés, l'organisme national de statistique peut évaluer les risques associés à la livraison en temps voulu des données administratives par le fournisseur, en fonction des critères de qualité requis.

IV. Conclusions et recommandations

70. Les données administratives peuvent être utilisées pour appliquer les différentes méthodes de recensement et pour soutenir toutes les étapes du processus de recensement, notamment la constitution d'une base d'adresses, le soutien des opérations sur le terrain, le dénombrement de la population, la collecte des variables de recensement, l'assurance qualité, la vérification et l'imputation, la modélisation et l'estimation. Leur utilisation peut permettre d'obtenir des statistiques plus fréquentes et plus actuelles sur la population, d'améliorer la précision et la fiabilité et de réduire considérablement les coûts et la charge de travail des répondants.

71. Cependant, il y a d'importants problèmes de qualité à évaluer et à surmonter avant qu'une source administrative puisse être utilisée dans un recensement. Le plus important de ces problèmes est que généralement, les données administratives n'ont pas été collectées pour les besoins d'un recensement. Par conséquent, l'organisme national de statistique peut avoir un contrôle limité sur les concepts et les définitions utilisés, la population cible, les procédures de collecte, de traitement et d'assurance qualité, ainsi que sur les méthodes, structures et systèmes utilisés en matière de données.

72. Les Directives ont défini des étapes d'évaluation de la qualité en fonction d'un certain nombre de critères de qualité, en y associant des outils et des indicateurs afin de guider l'utilisateur au cours du processus d'évaluation. Leur application devrait aider les lecteurs à prendre des décisions sur l'utilisation des données administratives dans le recensement, tout en favorisant un processus d'évaluation et d'amélioration permanentes. Tout au long des

Directives, un certain nombre de propositions et de recommandations ont été faites, qui sont résumées ci-dessous.

A. Recommandations

73. Les recommandations suivantes ont été faites :

i. L'organisme national de statistique doit *déterminer quelles sont les sources administratives* qui peuvent être utiles au recensement, en fonction de différents cas d'utilisation. Il est important d'*établir quels seraient les résultats attendus ou requis de l'utilisation de ces sources*, résultats en fonction desquels une évaluation de la pertinence pourra être effectuée. Il pourrait s'agir d'une amélioration de l'efficacité des opérations de recensement par la réduction des coûts et de la charge de travail des répondants, d'une amélioration de la qualité du recensement ou de la production de résultats nouveaux ou améliorés. Pour cette évaluation, il est essentiel de définir ce que la source administrative doit apporter en ce qui concerne la population cible et les mesures à effectuer auprès de cette population dans le cadre du recensement. Les Directives complètes, y compris les études de cas, donnent des exemples de la manière dont les données administratives ont été utilisées dans plusieurs pays différents.

ii. La *relation entre l'organisme national de statistique et le fournisseur de données administratives* est d'une importance capitale. Elle doit être soutenue par de solides mécanismes de communication, des accords écrits et une excellente compréhension des besoins des deux parties. Il doit également exister une base juridique convenue pour la fourniture et l'utilisation des données. Pour aider à établir cette relation et à garantir la fourniture de données, l'organisme national de statistique doit déterminer quels sont les domaines qui présentent un avantage pour le fournisseur. Il peut s'agir de mécanismes de retour d'information susceptibles d'aider le fournisseur à mieux comprendre les données dont il dispose, de collaborations dans des domaines d'intérêt commun ou d'une aide apportée au fournisseur (par l'utilisation de ses données dans le cadre du recensement) dans sa contribution à l'intérêt général. Bien entendu, le fait de signaler les éventuels problèmes de qualité des données présente l'avantage supplémentaire de faciliter l'amélioration continue de la qualité.

iii. *L'organisme national de statistique doit collaborer avec le fournisseur pour acquérir une connaissance approfondie de la source de données*. Cela doit se traduire par la création de métadonnées claires et détaillées sur la source administrative. Ces métadonnées constitueront une référence utile tant pour le recensement que pour toute autre statistique qui pourrait bénéficier de l'utilisation de la source. Les Directives complètes fournissent des détails sur les métadonnées qui doivent être collectées, ainsi que diverses références à la littérature pertinente.

iv. Étant donné que les données administratives ne sont généralement pas recueillies pour les besoins du recensement, il est important que l'organisme national de statistique *comprenne et évalue les différences entre les populations, les concepts, les définitions et les aspects temporels requis*. Plus généralement, il est essentiel de procéder à une évaluation approfondie de la cohérence et de la comparabilité de la source administrative, ainsi que de ses limites, en fonction des différents critères de qualité. Cela inclut la capacité de couplage de la source, s'il s'agit d'une exigence pour l'utilisation dans le recensement. Cette évaluation servira de base aux étapes du traitement, notamment le mappage et la dérivation, la vérification et l'imputation, ainsi que le couplage et l'intégration des sources (lorsque des décisions sont prises entre les sources et à l'intérieur de celles-ci sur la base de leur qualité).

v. L'organisme national de statistique doit *comprendre toutes les restrictions et difficultés liées à l'obtention d'une source administrative et à son intégration dans le recensement*. Il peut s'agir notamment des ressources et des coûts, des risques associés à la capacité du fournisseur de livrer à temps des données ayant la qualité requise et de la question de savoir si l'utilisation d'une source est acceptable pour le public et les utilisateurs des données de recensement. À cet égard, il existe d'importants choix

que l'organisme national de statistique doit prendre en considération. Plus précisément, *la valeur de la source administrative doit être évaluée en fonction de son utilité pour le recensement, par rapport aux efforts et aux risques liés à l'obtention et à l'utilisation des données.*

vi. L'organisme national de statistique n'a qu'un contrôle limité sur la collecte et le traitement des données administratives, qui peuvent être soumises au fil du temps à des changements dans la couverture de la population et dans les mesures effectuées auprès de celle-ci. Cela peut être dû, par exemple, à des changements d'ordre juridique ou des changements de politique, de procédure ou de système ayant une incidence sur les données et/ou leur fourniture. *L'organisme national de statistique doit donc évaluer et gérer un certain niveau de risque.* Il convient de gérer ce risque en travaillant avec le fournisseur de données sur les changements potentiels ou prévus, en faisant preuve de souplesse et en s'adaptant aux changements, ainsi qu'en réduisant dans la mesure du possible la dépendance à l'égard d'une seule source ou rubrique de données, que ce soit en utilisant d'autres sources de données ou en adaptant les processus ou les méthodes.

vii. *Il est important que le public et les utilisateurs de données comprennent comment et pourquoi les données administratives sont utilisées dans le recensement.* L'organisme national de statistique doit donc faire preuve de transparence au sujet de cette utilisation, en justifiant clairement les avantages par rapport aux risques et aux coûts éventuels (c'est-à-dire qu'il doit y avoir une forte proportionnalité). Cet objectif peut être atteint par *une bonne communication, y compris la publication des procédures et des politiques en place qui contribuent à l'efficacité de l'utilisation et de la protection des données.*

viii. L'inclusion de sources de données administratives dans la production du recensement doit être précédée d'une étude de faisabilité qui doit être suffisamment financée et permettre de valider le principe de cette intégration de données. Il est conseillé d'effectuer un certain nombre de tests (en utilisant des données réelles) bien avant le recensement principal afin de s'assurer que tous les problèmes imprévus sont répertoriés, en laissant suffisamment de temps pour corriger ou ajuster les méthodes, les processus ou les systèmes.

ix. *L'examen par des experts* (en collaboration avec les fournisseurs de données et les spécialistes du domaine) *et les comparaisons entre les sources et dans le temps sont importants pour repérer tout problème de qualité* d'une source ou d'un registre. L'utilisation d'enquêtes bien conçues (liées aux données administratives ou aux registres) peut être particulièrement importante pour déceler et corriger les erreurs de couverture et de mesure.

x. *L'organisme national de statistique doit enregistrer et publier les résultats de l'évaluation et de l'assurance de la qualité tout au long de la production du recensement*, y compris lors des étapes relatives aux données, aux processus et aux résultats. Cela permettra aux producteurs et aux utilisateurs du recensement de procéder à une évaluation et d'assurer un retour d'information, en favorisant ainsi un dialogue permanent. Il est important de veiller à ce que les utilisateurs comprennent les points forts et les limites et puissent aider à déterminer si le bon équilibre a été trouvé entre les différents critères de qualité.

xi. L'organisme national de statistique doit *élaborer son propre cadre et sa propre stratégie d'assurance qualité*, étayés par des procédures de documentation et de formation claires et détaillées. Les présentes Directives constituent une base utile à cet effet, de même que le matériel de référence et les études de cas qu'elles contiennent. La stratégie doit intégrer l'évaluation et l'amélioration continues des données administratives dans les plans et procédures du recensement. Les liens de communication entre l'organisme national de statistique, les utilisateurs et les fournisseurs de données doivent aussi en faire partie.

B. Domaines à approfondir

74. Les Directives ont mis l'accent sur l'évaluation de la qualité des sources administratives à utiliser dans les recensements, tout en donnant quelques informations sur les processus utilisés pour intégrer et transformer les données afin d'améliorer la qualité. En outre, la qualité des résultats des recensements qui utilisent des données administratives y est abordée brièvement. Toutefois, les Directives ne proposent pas de cadre plus large relatif à l'erreur totale ni de modèle permettant de déterminer comment l'erreur provenant de chaque source se traduit par une erreur dans les estimations finales du recensement, compte tenu des variations de qualité dues au traitement (qui peut réduire ou augmenter l'erreur).

75. L'élaboration d'un tel modèle qui tiendrait compte de toutes les sources d'erreur est en partie abordée par le cadre relatif à l'erreur totale adopté par Statistics New Zealand (Reid *et al.*, 2017). Ce cadre s'appuie sur l'extension par Li-Chun Zhang (Zhang 2012) du paradigme de l'erreur totale dans les enquêtes (voir Groves & Lyberg 2010 ; Biemer 2010). Il comporte trois phases couvrant : 1) une évaluation des sources uniques ; 2) une évaluation de l'ensemble intégré de données ; et 3) une évaluation de l'estimation et des résultats. Les travaux réalisés dans le cadre du projet ESSnet KOMUSO sur la qualité des statistiques issues de sources multiples (WPI Quality, 2019) proposent aussi un cadre utile pour évaluer la qualité des résultats statistiques obtenus à partir de sources multiples (données d'enquête et données administratives).

76. Ce domaine pourrait faire l'objet de travaux et d'une collaboration internationale plus poussés, qui mettraient l'accent sur la manière dont un tel cadre peut s'appliquer aux recensements. Il pourrait s'agir d'examiner comment un cadre ou un modèle d'erreur totale peut être élaboré et utilisé pour évaluer la qualité des résultats des recensements fondés sur des sources multiples. Il pourrait également s'agir de comprendre comment l'incidence (et l'incidence cumulée) de diverses erreurs commises au cours des différentes étapes du recensement peuvent éclairer les décisions concernant la meilleure conception statistique globale du recensement.

77. Enfin, les Directives ont mis l'accent sur l'évaluation des données administratives, mais il existe aussi des *sources de données commerciales qui offrent des possibilités d'utilisation pour améliorer ou renforcer les statistiques de recensement* (par exemple les données géospatiales et les données de téléphonie mobile). Les étapes, critères, outils et indicateurs relatifs à la qualité présentés dans les Directives sont, dans une large mesure, applicables à des sources autres que les données administratives. Il s'agit là aussi d'un domaine qui pourrait faire l'objet de travaux internationaux supplémentaires, qui porteraient en particulier sur la question de savoir si et comment les outils et les techniques d'évaluation de la qualité de ces sources en vue de leur utilisation dans le cadre du recensement diffèrent des outils et ses techniques qui sont présentés ici.
