

**Commission économique pour l'Europe****Conférence des statisticiens européens****Soixante-sixième réunion plénière**

Genève, 18-20 juin 2018

Point 4 b) de l'ordre du jour provisoire

Intégration des données pour mesurer les migrations**Document d'orientation sur l'intégration des données
pour mesurer les migrations****Note de l'Équipe spéciale de l'intégration des données
pour mesurer les migrations***Résumé*

Le présent document contient une version abrégée du *document d'orientation sur l'intégration des données pour mesurer les migrations*, qui a été établi par une équipe spéciale composée de l'Espagne (présidence), de l'Autriche, des États-Unis, de l'Italie, de la Nouvelle-Zélande, du Royaume-Uni, de la Suisse, de l'Office statistique de l'Union européenne (Eurostat) et de la Commission économique pour l'Europe (CEE).

La présente version abrégée du document d'orientation est établie à des fins de traduction et ne comprend pas les études de cas nationales. Le document d'orientation comprend les éléments suivants : un examen de la définition de l'intégration des données se référant en particulier aux statistiques des migrations internationales (chap. II) ; un examen des résultats de l'enquête sur les pratiques en matière d'intégration des données (chap. III) ; un résumé des résultats de 13 études de cas nationales (chap. IV) ; une liste de métadonnées et une réflexion à ce sujet (chap. V) ; les conclusions (chap. VI) ; et un certain nombre de recommandations générales à l'intention des pays (chap. VII).

La version complète du document d'orientation (comprenant les études de cas nationales) a été adressée par voie électronique à tous les membres de la Conférence des statisticiens européens pour consultation. Elle est disponible à l'adresse <http://www.unece.org/index.php?id=47411>. Si la consultation donne des résultats positifs, la Conférence sera invitée à approuver ce document d'orientation à sa réunion plénière.



Table des matières

	<i>Paragraphes</i>	<i>Page</i>
I. Introduction	1–10	3
II. Définition de « l'intégration des données ».....	11–31	5
A. Généralités	11–23	5
B. Définition de travail	24–31	7
III. Enquête sur les pratiques en matière d'intégration de données	32–47	8
A. Aperçu des principaux résultats de l'enquête	35–40	9
B. Recommandations.....	41–42	9
C. Synthèse des résultats	43–47	10
IV. Études de cas	48–49	11
A. Synthèse des études de cas.....	50–70	11
V. Métadonnées	71–77	14
VI. Conclusion	78–88	16
VII. Recommandations	89–98	17
A. Améliorer l'accès aux données administratives pour les organismes nationaux de statistique.....	90	17
B. Utiliser les données administratives pour établir des statistiques migratoires	91–92	17
C. Combiner les données provenant de différentes sources au moyen des « signes de présence ».....	93	18
D. Veiller à la qualité des données intégrées	94	18
E. Présenter de manière transparente les méthodes d'intégration des données et élaborer des normes	95	18
F. Favoriser la comparaison internationale et l'échange de données sur les migrations	96–98	18
VIII. Travaux futurs	99–103	19
A. Le potentiel des mégadonnées	100	19
B. Données longitudinales.....	101–103	19
Références		20

I. Introduction

1. L'intégration des données est un sujet qui a suscité un vif intérêt ces dernières années, d'autant plus que les besoins en données se sont accrus dans des contextes budgétaires plus rigoureux, associés à une augmentation de la charge que représente la communication de données et à des préoccupations en matière de protection de la vie privée. Le Groupe de haut niveau de la CEE sur la modernisation des statistiques officielles (le Groupe de haut niveau) a mis en œuvre un projet sur l'intégration des données au cours de la période 2016-2017. Ce projet comprenait des expériences d'intégration des données dans différents domaines et a conduit à l'élaboration d'un document d'orientation pratique en ligne sur l'intégration des données pour les statistiques officielles¹. Eurostat a appuyé les activités de recherche sur les méthodes permettant d'améliorer l'intégration des données, telles que le projet ESSnet dans le domaine de l'intégration des données provenant d'enquêtes et des données tirées de fichiers administratifs². Ce projet a constitué une première tentative visant à créer une base méthodologique commune pour l'intégration de différentes sources de données.

2. Toutefois, de nombreux pays s'intéressent à l'intégration des données plus particulièrement aux fins de la mesure des migrations internationales. Compte tenu des difficultés que pose la collecte de statistiques sur les migrations, il est souvent utile de recueillir des données auprès de plusieurs sources différentes. L'intégration des données peut réduire les problèmes de couverture ou d'exactitude qui peuvent concerner l'ensemble des données sur les effectifs et les flux migratoires, et elle peut accroître la qualité des données relatives aux migrations en ajoutant des dimensions sociodémographiques et économiques aux données existantes.

3. Au niveau européen, l'importance de l'intégration des données pour améliorer les statistiques sur les migrations a été réaffirmée à l'occasion de la Conférence annuelle de 2017 des directeurs généraux des instituts nationaux de statistique, qui comprenait une session consacrée au thème « Mouvements de population et questions liées à l'intégration – Statistiques sur les migrations »³. Le « Mémoire de Budapest » examiné à cette conférence et approuvé par le Comité du Système statistique européen prévoit notamment les mesures concrètes suivantes : « *Apporter un appui au recensement, à l'évaluation et à l'adoption de nouvelles méthodes de travail et sources de données, en particulier l'utilisation accrue à des fins statistiques de sources de données administratives de qualité appropriée assurée par une évaluation continue de la qualité (telles que des registres individuels ou des données couplées provenant de plusieurs sources administratives ou combinées avec des sources provenant d'enquête), et les possibilités offertes par de nouvelles sources de données (par exemple, les mégadonnées)* ».

4. Les méthodes d'intégration de différentes sources de données administratives d'un pays (par exemple, les registres de la population, les bases de données sur la santé, le travail, la fiscalité ou l'éducation) utilisées pour obtenir les informations manquantes dans diverses sources (par exemple, pour mesurer l'émigration des personnes qui ne se sont pas radiées des registres) se sont avérées efficaces dans certains pays. Dans d'autres cas, des sources de données non administratives peuvent fournir des informations manquantes dans les sources administratives (par exemple, l'utilisation d'enquêtes auprès des ménages pour collecter des informations qui ne figurent pas dans les registres). On envisage également dans certains pays d'utiliser les « mégadonnées » des compagnies de services publics et de téléphonie comme complément d'information. L'intégration des données pourrait également être utilisée pour harmoniser des chiffres différents sur les migrations provenant de différentes sources, en particulier pour ce qui concerne les estimations des groupes de migrants « difficiles à dénombrer », tels que les migrants en situation irrégulière ou les émigrants.

¹ <https://statswiki.unece.org/display/DI/Data+Integration+Home>.

² https://ec.europa.eu/eurostat/cros/content/data-integration_en.

³ 103^e Conférence des directeurs généraux des instituts nationaux de statistique (Budapest, 20 et 21 septembre 2017). Les documents et les exposés peuvent être consultés à l'adresse : <http://www.ksh.hu/dgins2017/presentations.html>.

5. Notant l'importance et l'actualité de ce sujet, les participants à la réunion de travail sur les statistiques des migrations tenue par la CEE et Eurostat en 2014 ont échangé des vues sur l'utilisation de différentes sources de données administratives pour mesurer les migrations internationales. Les débats ont porté sur les possibilités et les défis liés à l'utilisation de données administratives, compte tenu en particulier des différents niveaux de développement des systèmes de données administratives dans la région. Les moyens d'améliorer la coopération entre les services de migration, les organismes de statistique, les services chargés des registres et d'autres producteurs de données administratives au niveau national ont été examinés, ainsi que les méthodes permettant d'intégrer diverses sources de données administratives pour améliorer la mesure des migrations. Les participants ont également souligné la nécessité d'une orientation et de conseils pratiques sur la production de métadonnées afin de faciliter les comparaisons entre les estimations sur les migrations effectuées par les différents pays.

6. À l'issue de cette réunion de travail, les participants ont formulé une recommandation visant à ce que soient entrepris des travaux méthodologiques supplémentaires sur les thèmes de l'intégration de multiples sources de données pour mesurer les migrations internationales, notamment les sources de données se trouvant dans un pays ou étant réparties entre différents pays, et de l'échange de bonnes pratiques en matière de communication entre les organismes nationaux de statistique et les producteurs de données administratives.

7. Comme suite à ces débats, le Bureau de la Conférence des statisticiens européens⁴ a créé en octobre 2015 l'Équipe spéciale de l'intégration des données pour mesurer les migrations et l'a chargée d'élaborer un ensemble de lignes directrices et règles de bonnes pratiques concernant l'intégration de différentes sources de données pour améliorer la mesure de l'immigration, de l'émigration et de la migration nette. La publication expose les résultats de ce travail.

8. La publication donne un aperçu général de ce sujet ainsi qu'une vue d'ensemble des types d'intégration des données qui sont déjà mis en œuvre dans divers pays – qu'ils soient diffusés dans le cadre de publications régulières de données ou d'études pilotes. Les principes des meilleures pratiques établis sur la base de cette vue d'ensemble sont également présentés afin de servir de lignes directrices concernant les moyens d'améliorer l'intégration des données pour mesurer les migrations dans différents pays. Les données d'expérience nationales sont présentées dans la publication sur la base d'une enquête réalisée auprès des fournisseurs de données sur les migrations dans près de 50 pays ainsi que d'études de cas plus détaillées concernant plusieurs pays.

9. La publication traite de l'intégration des données en examinant à la fois *l'intégration des macrodonnées* (la comparaison/modélisation statistique fondée sur des données qui sont des agrégats (statistiques) de données enregistrées au niveau individuel) et *l'intégration des microdonnées* (l'intégration de données fondée sur les couplages/appariements statistiques de fichiers enregistrés au niveau individuel). Les différents niveaux de chevauchement des variables et/ou des individus entre différentes sources sont également analysés.

10. De plus amples détails sur les notions susmentionnées sont fournis dans le cadre des considérations théoriques et de la définition de l'intégration des données figurant au chapitre II. Le chapitre III présente une analyse des résultats de l'enquête réalisée auprès des fournisseurs de données sur les migrations. Des exemples concrets d'intégration des

⁴ La Conférence des statisticiens européens réunit les représentants des organismes nationaux de statistique de la région de la CEE (pour la liste des États membres de la CEE, voir https://www.unece.org/oes/nutshell/member_states_representatives.html) ainsi que de l'Australie, du Brésil, du Chili, de la Chine, de la Colombie, du Japon, du Mexique, de la Mongolie, de la Nouvelle-Zélande et de la République de Corée. Participent également aux travaux de la Conférence les principales organisations internationales de la région de la CEE actives dans le domaine de la statistique, comme l'Office statistique de la Commission européenne (Eurostat), l'Organisation de coopération et de développement économiques (OCDE), le Comité inter-États de statistique de la Communauté d'États indépendants (CIS-STAT), l'Organisation internationale du Travail (OIT), le Fonds monétaire international (FMI), l'Organisation mondiale du commerce (OMC) et la Banque mondiale.

données dans le domaine des migrations sont illustrés par des études de cas approfondies concernant un certain nombre de pays. Le résumé des résultats des études de cas est présenté au chapitre IV (les études de cas sont présentées dans la version complète de la publication, disponible en anglais seulement). Une proposition relative aux métadonnées à examiner en ce qui concerne les statistiques des migrations internationales est présentée au chapitre V. Enfin, les conclusions et les recommandations générales fondées sur l'enquête et les études de cas sont présentées au chapitre VI.

II. Définition de « l'intégration des données »

A. Généralités

11. La mesure des migrations est depuis longtemps une tâche particulièrement difficile. Bien que des efforts considérables aient été accomplis récemment dans le domaine des statistiques universitaires et officielles afin d'améliorer la qualité des statistiques des migrations, il existe encore des marges d'incertitude non négligeables quant à leur exactitude. Plus récemment, dans un contexte statistique plus large, l'attention a été appelée sur les possibilités offertes par l'« intégration des données » en tant que moyen supplémentaire susceptible d'améliorer les statistiques (par exemple, Eurostat, 2009, 2013 ; ESSnet, 2008, 2011, 2014 ; Groupe de haut niveau de la CEE sur la modernisation des statistiques officielles, 2017).

12. L'« intégration des données » est souvent mentionnée comme l'une des approches les plus efficaces pour enrichir et améliorer les informations statistiques. Au cours des dernières années, divers projets et activités ont porté sur l'intégration des données et ont donné lieu à l'adoption de définitions différentes. Par exemple, selon le Modèle générique du processus de production statistique, l'intégration des données est une activité qui consiste à intégrer des données provenant d'une ou de plusieurs sources dans le cadre du processus de production statistique. Dans le projet du Groupe de haut niveau susmentionné, l'intégration des données est définie comme « l'activité consistant à associer dans un ensemble de données au moins deux sources de données différentes »⁵. Mais il n'existe pas, à notre connaissance, de définition de l'intégration des données internationalement reconnue par la communauté des statisticiens. Diverses caractéristiques peuvent aider à caractériser l'intégration des données, telles que le nombre et la typologie des sources de données, la méthodologie, l'actualité, la charge que représente la communication des données et, surtout, l'objectif de l'intégration.

13. Intuitivement, on comprend que l'intégration nécessite au moins deux éléments en entrée. Ces éléments sont des ensembles de données, c'est-à-dire des informations organisées établies à partir de certaines sources de données au moyen d'une opération statistique. Dans le cas des migrations, les ensembles de données peuvent être classés en quatre catégories : (ensembles de données établis à partir des) enquêtes statistiques (y compris les enquêtes exhaustives, c'est-à-dire les recensements), registres administratifs, mégadonnées et informations géographiques. Ces types de données peuvent être intégrés selon toutes les combinaisons possibles et ce processus peut être répété, générant des ensembles de données de nature mixte qui peuvent à leur tour être intégrés dans d'autres ensembles de données.

14. La façon dont les ensembles de données sont intégrés dépend principalement de la question de savoir si les informations qu'ils contiennent sont des microdonnées ou des macrodonnées et, pour les données enregistrées au niveau individuel, de la disponibilité d'un identifiant commun. Pour l'intégration au niveau des microdonnées, cette dernière caractéristique oriente le choix entre l'appariement statistique et le couplage de fichiers, cette deuxième option étant généralement appliquée lorsqu'un identifiant est disponible dans tous les ensembles de données intégrés. En ce qui concerne l'intégration au niveau des macrodonnées, c'est-à-dire l'intégration de données agrégées à partir de données individuelles, le processus devrait se dérouler en deux étapes : premièrement, éliminer les

⁵ <https://statswiki.unece.org/pages/viewpage.action?pageId=169018059>.

différences existant entre les concepts et les définitions opérationnelles ; deuxièmement, harmoniser (« équilibrer ») les données, en utilisant diverses méthodes statistiques ou simplement une opinion d'expert.

15. Dans le cas des données sur les migrations, l'identifiant individuel le plus courant est le numéro d'identification personnel (PIN). En fait, l'intégration au niveau des microdonnées est préférable pour l'établissement des registres de la population, dont les statistiques migratoires sont souvent un sous-produit. Il convient de noter que le PIN n'est pas le seul identifiant commun possible, car différents codes d'identification peuvent être utilisés pour relier les données relatives aux agrégats infranationaux, aux entités administratives, aux fichiers, etc.

16. Les sources de données utilisées pour établir un registre de la population peuvent être nombreuses et, dans certains pays, d'autres sources de données ont été intégrées au fil du temps, ce qui a permis d'enrichir les informations statistiques sur la population et, indirectement, sur les migrations. Cela peut aussi être le cas pour les registres de la population étrangère, ou registres des étrangers, même si cela se fait généralement dans une moindre mesure. L'intégration des données peut donc être considérée comme un processus en expansion, dans lequel les premiers ensembles de données intégrés sont complétés par d'autres grâce aux enseignements tirés, au perfectionnement des méthodes d'intégration et à l'accès à de nouvelles sources de données.

17. Une autre caractéristique spécifique à la mesure des migrations tient au fait que, étant donné que ce phénomène concerne à la fois deux pays (le pays de destination et le pays d'origine, ou le pays d'accueil et le pays de naissance, ou toute autre variable d'identification des migrants), d'autres sources de données que les sources nationales peuvent être utilisées. L'exploitation de cette caractéristique internationale intrinsèque se limite principalement à l'échange de données agrégées, étant donné que des préoccupations importantes en matière de protection de la vie privée et de sécurité nationale peuvent être opposées par les pays à toute demande d'échange international de microdonnées sur les migrations. On trouve des exemples de l'approche susmentionnée dans les pays nordiques, où l'échange international de microdonnées est utilisé pour améliorer la qualité des registres de la population.

18. L'ajout de nouvelles sources de données peut entraîner des changements dans le degré d'actualité de la production statistique, c'est-à-dire modifier les délais entre la période de référence et le moment où les résultats sont disponibles. Alors qu'en règle générale l'obtention d'un produit statistique établi à partir de données intégrées est plus rapide lorsque la source de données comporte des délais plus larges, la modélisation statistique peut en fait inverser cette règle en offrant la possibilité de publier plus rapidement des estimations. Il est probable que le degré moins élevé d'actualité d'une mesure statistique ne sera accepté que si l'intégration des données permet d'améliorer de manière significative la couverture ou l'exactitude de cette mesure.

19. La réduction de la charge que représente la communication de données est parfois mentionnée comme l'un des avantages offerts par l'intégration des données. Jusqu'à présent, l'utilisation la plus efficace d'autres sources de données pour les statistiques des migrations a été obtenue en remplaçant une source de données par une autre source faisant peser une moindre charge sur les répondants/fournisseurs de données ou en améliorant l'opération statistique appliquée à la source de données. D'un point de vue théorique, cela peut être considéré comme une amélioration en termes d'efficacité plutôt que de disponibilité des données, et ce n'est donc pas exactement un avantage propre à l'intégration des données.

20. Les données provenant d'autres sources peuvent également être utilisées pour valider le produit statistique d'une seule source de données officielles. Dans ce cas, on peut considérer que ce processus a un rôle de *comparaison* plutôt que d'*intégration*. Cela peut également constituer la première étape d'un processus d'intégration progressive de ces sources de données, en particulier lorsque les sources supplémentaires ne sont pas conformes aux exigences des statistiques officielles.

21. Un autre cas se présente lorsque chaque source de données couvre un sous-groupe particulier de la population (migrants). Le regroupement de ces données pour obtenir un

produit statistique global pourrait être qualifié de « compilation » plutôt que d'« intégration » au sens propre, puisqu'il n'y a pas de fusion réelle au niveau de la donnée enregistrée individuelle (individus ou entités agrégées).

22. Il ressort des considérations qui précèdent que la caractéristique déterminante de l'intégration des données est sans aucun doute l'utilisation de plusieurs ensembles de données, mais que cela dépend également de la manière dont ces ensembles sont utilisés conjointement. En fait, certaines opérations statistiques, comme la compilation et la comparaison de données, sont très proches de l'intégration des données. En ce qui concerne les autres caractéristiques, le niveau de détail des données est important pour déterminer la méthode à appliquer, plutôt que pour statuer sur une activité d'intégration des données. La rapidité d'obtention des statistiques ne peut pas non plus être utilisée comme critère de choix, étant donné que l'intégration des données peut en principe aussi bien allonger que raccourcir les délais d'obtention des résultats, et qu'on peut également obtenir une amélioration en termes de rapidité en appliquant d'autres opérations statistiques que l'intégration. De même, il n'est pas pertinent de se référer de manière générale à l'amélioration de la qualité des produits statistiques (notamment à la réduction de la charge que représente la communication des données) car toute opération statistique devrait en principe viser un tel effet. À l'évidence, cela ne signifie pas que l'intégration des données n'a pas ces effets positifs, mais seulement que ces critères ne sont pas pertinents pour définir les caractéristiques distinctives de l'intégration des données.

23. Le résultat d'une activité d'intégration des données devrait être un ensemble de données « enrichi » ou « de meilleure qualité ». Si l'on envisage la question sous la forme d'une matrice de données génériques $n \times p$, dans laquelle n correspond aux données enregistrées et p aux variables, un tel enrichissement pourrait être obtenu dans les deux sens, par une amélioration de la couverture (à savoir en augmentant la valeur de n) et/ou des informations sur les mêmes données enregistrées (à savoir en ajoutant de nouvelles variables à l'ensemble de facteurs p). Quant à l'amélioration de la qualité, elle ne se traduirait pas par des modifications de la taille de l'ensemble de données qui en résulterait, mais plutôt de son contenu (par exemple dans les pondérations d'une enquête par sondage).

B. Définition de travail

24. Le milieu informatique (SDMX, 2009) donne la définition suivante de l'intégration des données : « *Processus consistant à associer des données provenant de deux ou de plusieurs sources pour élaborer des produits statistiques* ». Dans la même référence, il est également précisé que « *l'intégration des données peut se faire au niveau des microdonnées, où elle est souvent qualifiée d'appariement, ou au niveau des macrodonnées* ».

25. La définition ci-dessus repose sur les critères des entrées (« *deux ou plusieurs sources* ») et de l'objectif (« *élaborer des produits statistiques* »). Ce deuxième critère est assez générique car un produit statistique peut être considéré comme la finalité de tout processus statistique et ne requiert pas nécessairement l'intégration des données.

26. Aux fins du présent document, la définition de travail suivante est proposée : « *L'intégration des données est une activité statistique portant sur deux ou plusieurs ensembles de données et aboutissant à un ensemble de données unique élargi et/ou de meilleure qualité* ».

27. L'intégration des données peut être traitée à deux niveaux principaux d'agrégation, celui des microdonnées ou celui des macrodonnées⁶, l'intégration étant définie comme suit :

a) « *Intégration de microdonnées* » : intégration de données fondée sur le couplage de fichiers/l'appariement statistique de données enregistrées au niveau individuel à l'aide de variables d'identification clefs ;

⁶ Dans certains contextes, l'expression « micro-macrodonnées » est utilisée lorsque des microdonnées et des macrodonnées sont associées. Cette approche peut être considérée ici comme un cas d'intégration de macrodonnées.

b) « Intégration de macrodonnées » : combinaison de données fondée sur des agrégats (statistiques) de données enregistrées au niveau individuel.

28. La référence à une « activité statistique » générique laisse ouverte la possibilité de recourir à n'importe quelle méthode, que ce soit au niveau des macrodonnées ou des microdonnées, y compris les avis d'experts. Elle couvre également les travaux portant sur les différences conceptuelles, qui constituent une étape primordiale dans tout processus statistique. La référence explicite à l'appariement statistique et/ou au couplage de fichiers, bien qu'elle soit propre à l'intégration des données, aurait exclu l'intégration au niveau des macrodonnées.

29. Les termes « *deux ou plusieurs ensembles de données* » soulignent le caractère multidimensionnel de l'intégration des données. Ils indiquent en outre que, bien que l'intégration aboutisse à un résultat unique, il s'agit d'une activité qui doit être répétée chaque fois que l'on souhaite obtenir un tel résultat. En d'autres termes, pour une production régulière de statistiques à partir de plus d'un ensemble de données, l'intégration n'est pas une activité statistique occasionnelle qui ne serait effectuée qu'une seule fois. Le résultat « intégré » est généré chaque fois qu'une entrée de données plus récentes est disponible⁷.

30. L'utilisation des termes « ensembles de données » au lieu de « sources de données » indique que les données faisant l'objet de l'intégration sont des données réelles, et non les sources de données à partir desquelles elles sont obtenues. Par exemple, lorsque l'on regroupe deux enquêtes par sondage existantes en une seule enquête globale en modifiant les questionnaires, la conception de l'échantillonnage, etc., cela n'est pas considéré dans ce contexte comme une intégration de « données ». Cela corrobore également l'avis exprimé plus haut selon lequel l'intégration ne transforme pas les sources de données utilisées en nouvelles sources de données mixtes : ce sont les données obtenues à partir de ces sources qui constituent le cœur de l'activité. L'opération d'intégration est donc susceptible d'être aussi systématique que la production de statistiques à partir des données auxquelles elle est appliquée.

31. Le résultat de l'intégration des données est d'abord et avant tout un ensemble de données « unique » : un nouvel ensemble de données dans lequel toutes les informations provenant des ensembles de *données d'entrée* sont réorganisées de manière harmonisée. Il ne s'agit pas simplement du regroupement de plusieurs ensembles de données, car les redondances, les différences conceptuelles et tout autre facteur de biais sont censés être traités correctement. Il s'agit d'un ensemble de données « élargi », c'est-à-dire d'un nouvel ensemble structuré de données dont les dimensions ne peuvent être inférieures aux dimensions correspondantes les plus grandes des ensembles de données d'entrée individuels, et/ou d'un ensemble de données « de meilleure qualité », dont les éléments ont été modifiés en raison de l'intégration des données, ce qui a entraîné un gain (éventuellement mesurable) en qualité statistique.

III. Enquête sur les pratiques en matière d'intégration de données

32. Pour recueillir les informations dont elle avait besoin, l'Équipe spéciale a réalisé une enquête sur les pratiques en matière d'intégration de données. Un questionnaire spécifique a été conçu pour recueillir des informations sur ce qui se faisait actuellement concernant l'utilisation de sources de données combinées ou intégrées pour mesurer les flux d'immigration et d'émigration et produire des statistiques relatives à la population migrante et à la population d'origine étrangère.

⁷ La mise en évidence de différences conceptuelles dans le contexte de l'intégration des données constitue peut-être une exception à cette « règle de la répétition », car une telle activité ne devrait être effectuée qu'une seule fois, à moins que le cadre conceptuel des données d'entrée ne change au fil du temps.

33. Le questionnaire a été distribué en septembre 2016 aux organismes nationaux de statistique des pays membres de la CEE⁸. Cinquante-six pays ont répondu aux parties du questionnaire qui concernaient la façon dont ils collectaient les données.

34. Les résultats de l'enquête ont été analysés entre la fin de 2016 et le premier trimestre de 2017. Les principaux d'entre eux sont résumés dans les paragraphes suivants.

A. Aperçu des principaux résultats de l'enquête

35. Il est essentiel d'intégrer des données sous une forme ou sous une autre pour pouvoir produire des statistiques relatives aux flux migratoires internationaux. Lorsque cela est possible, l'intégration est réalisée au niveau des données individuelles, mais dans l'ensemble, une telle solution n'est envisageable que dans les pays qui disposent de registres de la population complets où sont comptabilisés les départs de nationaux et les arrivées d'étrangers.

36. Dans de nombreux pays, l'événement migratoire proprement dit n'est pas consigné par une source unique et les statistiques sur les flux migratoires ne peuvent être obtenues qu'en fusionnant les éléments qui contribuent à faire évoluer la population résidente du pays à la faveur des mouvements migratoires.

37. Il est possible de produire des statistiques relatives aux mouvements migratoires internationaux à partir d'une source unique lorsque les données relatives à l'événement migratoire (franchissement de la frontière) peuvent être recueillies et gérées par voie administrative ou lorsqu'il existe un système reposant sur un registre de la population complet. Toutefois, l'assurance de la qualité, puis la ventilation de ces statistiques peuvent nécessiter l'intégration de sources de données auxiliaires, ce qui peut se faire simplement en ajustant les mouvements migratoires après comparaison avec les statistiques disponibles dans d'autres pays (statistiques miroirs). Il peut être nécessaire de regrouper les individus au sein d'une source unique pour réaliser des observations longitudinales des mouvements aux frontières et opérer ainsi une classification par statut migratoire.

38. L'enquête a mis en évidence une grande diversité dans l'utilisation de l'intégration de données. Dans certains pays, il n'est parfois pas possible de relier au niveau des unités individuelles les registres des nationaux à une source de données administratives distincte sur les ressortissants étrangers. En de tels cas, on peut utiliser le registre de la population pour réaliser des vérifications croisées et des imputations des données manquantes.

39. Les recensements ou les registres de la population sont des sources fréquemment utilisées pour produire des statistiques concernant la population d'origine étrangère. Les recensements peuvent servir à établir une population de référence, après quoi l'intégration avec une source de données administratives permet d'affiner les statistiques relatives à la population d'origine étrangère.

40. Certains des pays qui produisent leurs statistiques sur les migrations à partir de sources multiples ont indiqué qu'il était nécessaire de procéder à une évaluation de l'ordre de priorité des différentes sources en fonction de leur actualité, de leur exhaustivité, de leur fiabilité, de leur cohérence et d'autres critères de qualité.

B. Recommandations

41. Il est probable que l'intégration de données provenant de sources multiples au niveau des unités individuelles tendra à se renforcer au fur et à mesure que les pays progresseront dans la normalisation des registres de statistiques sociodémographiques. L'évaluation de la qualité des données intégrées est appelée à faire l'objet d'une attention accrue dans le contexte de la production de statistiques relatives aux migrations internationales.

⁸ Outre les 56 pays membres de la CEE, le questionnaire a également été envoyé aux pays qui participent régulièrement aux activités de la Conférence des statisticiens européens, notamment l'Australie, le Chili, la Colombie, le Japon, le Mexique et la Nouvelle-Zélande.

42. Il est de plus en plus important que les organismes de statistique présentent de manière transparente les règles et les méthodes qu'ils appliquent pour mener à bien les processus d'intégration de données qu'ils mettent en œuvre pour produire des statistiques relatives aux mouvements migratoires internationaux et d'autres formes de statistiques concernant les migrations internationales. Cela étendra en effet la possibilité d'élaborer de nouvelles normes méthodologiques d'intégration de données applicables par les différents organismes nationaux de statistique. Chaque pays disposera toujours de ses propres pratiques en matière de collecte et d'intégration des données, mais il est probable que davantage de lieux de rencontre ouverts se créeront pour examiner les règles et méthodes employées pour combiner les sources utilisées pour produire des statistiques sur les migrations.

C. Synthèse des résultats

43. Presque tous les pays qui ont répondu au questionnaire produisaient des statistiques sur les mouvements migratoires internationaux et plus de la moitié d'entre eux combinaient plusieurs sources de données. Certains pays utilisaient un registre centralisant toutes les inscriptions et toutes les radiations de résidents, qu'ils intégraient au niveau des unités individuelles avec les données administratives relatives aux mouvements d'immigration et d'émigration et aux décès, ce qui leur permettait de produire des statistiques plus précises concernant les flux migratoires. D'autres intégraient les registres de la population et les données des enquêtes ou d'autres ensembles de données administratives générales relatives aux étrangers.

44. Dans l'ensemble, environ un tiers des pays qui ont répondu au questionnaire (35 réponses au total) ont indiqué qu'ils intégraient les sources de données au niveau des macrodonnées en utilisant des modèles statistiques ou d'autres méthodes de combinaison d'agrégats, et une proportion légèrement inférieure de pays intégraient les données au niveau individuel. Il convient toutefois de noter que seuls quelques pays s'appuyaient sur d'autres sources pour réaliser des vérifications croisées et obtenir des informations complémentaires. Environ la moitié des pays ont fait état d'un chevauchement partiel des unités d'observation issues des sources de données intégrées. Ils étaient moins nombreux à indiquer que les unités d'observation obtenues à partir de sources combinées étaient identiques et seuls quelques-uns ont déclaré qu'ils utilisaient des unités d'observation qui s'excluaient mutuellement.

45. Moins de la moitié des pays établissaient des statistiques sur les mouvements migratoires internationaux à partir d'une source unique. Dans bien des cas, les statistiques de ce type peuvent être produites à partir d'une source unique lorsqu'il est possible de tenir un registre administratif de tous les franchissements de frontière ou d'observer les flux migratoires à partir d'un registre de la population. Une autre solution, moins fréquente, consiste à utiliser le recensement de la population pour estimer indirectement les flux migratoires en se basant sur les mesures des effectifs de la population.

46. Environ 90 % des pays qui ont répondu au questionnaire produisaient des statistiques ou collectaient des données concernant la population d'origine étrangère. Un peu plus de la moitié de ces pays s'appuyaient sur plus d'une source de données qu'ils intégraient au niveau des microdonnées ou des macrodonnées. Les principales sources utilisées, que ce soit en tant que source unique ou de façon intégrée, étaient un recensement quinquennal de la population, un registre de la population, les enquêtes ou les données administratives. Le principal objectif cité concernant l'intégration des données était la production de statistiques sur la population d'origine étrangère.

47. Environ 60 % des pays qui ont répondu au questionnaire ont indiqué qu'ils ne produisaient pas d'autres statistiques concernant les migrants ou la population d'origine étrangère. Par conséquent, seuls quelques pays ont indiqué que l'intégration des sources de données leur permettait de produire des statistiques concernant, par exemple, les demandeurs d'asile, les demandes d'asile acceptées, les permis de séjour, les détenteurs de permis de travail, les travailleurs étrangers en situation irrégulière, les caractéristiques socioéconomiques des migrants et les raisons qui les avaient conduits à s'installer dans le pays.

IV. Études de cas

48. S'appuyant sur les informations recueillies dans le cadre de l'enquête sur les pratiques en matière d'intégration des données, l'Équipe spéciale a constaté qu'un certain nombre de pays disposaient d'une expérience intéressante en matière d'intégration des données. Ces pays ont été invités à exposer brièvement par écrit les principales sources de données sur les migrations, les raisons pour lesquelles l'intégration des données avait été jugée nécessaire, les méthodes employées et les avantages obtenus.

49. On trouvera les textes présentés par les pays concernés dans la version intégrale du présent document (disponible en anglais seulement). Ces textes forment un recueil d'exemples pratiques intéressants d'intégration de données sur les migrations.

A. Synthèse des études de cas

50. Les études de cas montrent qu'il existe un large éventail de situations très diverses en ce qui concerne les sources permettant de produire des statistiques sur les migrations, les limites de ces sources et les démarches adoptées aux fins de l'intégration des données. Les exemples tirés des études de cas ne doivent pas être considérés comme pleinement représentatifs des démarches élaborées par les différents pays pour intégrer les données. En outre, il convient d'évaluer les expériences des différents pays en gardant à l'esprit le fait que les contextes nationaux peuvent varier très sensiblement d'un pays à l'autre sur un certain nombre d'aspects importants et qu'une démarche d'intégration de données qui aura fait ses preuves dans un pays ne sera pas nécessairement la meilleure des solutions possibles pour les autres. Cela étant, l'analyse des études de cas permet de tirer certains enseignements susceptibles d'intéresser d'autres pays.

51. En ce qui concerne *les pays qui disposent d'un registre de la population, celui-ci constitue, dans plusieurs cas (Autriche, Espagne, Italie et Lettonie), la principale source de données sur les migrations*. Le principal objectif de l'intégration des données est alors d'améliorer la qualité des données. Les données provenant d'autres registres peuvent être utilisées, notamment, pour réaliser les ajustements nécessaires afin de prendre en compte les radiations manquantes ou les décès. Dans plusieurs pays, on a également recours à l'intégration des données pour obtenir des informations supplémentaires sur les variables qui ne sont pas fournies par les sources primaires.

52. Plusieurs pays recourent à des pratiques basées sur les « *signes de présence* » (appelés « signes de vie » dans certains pays) des individus dans les divers registres. En *Autriche*, par exemple, les signes de présence dans les autres registres sont analysés de façon à estimer les personnes qui, tout en figurant dans le registre de la population, ne vivent plus dans le pays. Les données de la Sécurité sociale autrichienne sont utilisées pour détecter les personnes décédées et les données provenant d'autres sources administratives servent à opérer les ajustements pour les radiations manquantes.

53. En *Italie*, où seules les personnes qui résident légalement dans le pays peuvent figurer dans le registre de la population, l'analyse des signes de présence dans les diverses sources permet d'améliorer les estimations des migrations internationales en donnant la possibilité de détecter les personnes qui, tout en figurant dans le registre de la population, ont probablement émigré, ainsi que les personnes immigrées qui ne sont pas inscrites dans le registre de la population.

54. En *Lettonie*, les signes de présence sont analysés à partir de plusieurs registres, tels que les registres fiscaux et ceux de l'assurance sociale, de l'éducation, de la santé et de l'emploi. Les estimations relatives à l'émigration sont également produites à l'aide de statistiques miroirs des pays de destination. De plus, la qualité des statistiques sur les migrations est évaluée sur la base des données issues des enquêtes sur les ménages, notamment l'enquête sur les forces de travail et l'enquête sur les revenus et les conditions de vie de l'Union européenne.

55. En *Espagne*, le registre national de la population (« *padrón* ») est administré par l'Institut national de la statistique (INE) et toutes les personnes vivant dans le pays doivent

être enregistrées, quelle que soit leur situation juridique. Dans le cadre du recensement de 2011, le registre de la population a été croisé avec de nombreux autres registres administratifs de façon à pouvoir détecter les personnes dont la présence dans le pays était incertaine et les personnes décédées. S'agissant des entrées incertaines, les techniques d'intégration des données combinant des sources administratives et une enquête par sondage réalisée dans le cadre du recensement ont permis d'estimer le nombre de personnes figurant dans le registre, mais ne résidant pas effectivement dans le pays.

56. Dans certains pays (Hongrie, Israël, Pays-Bas et Suisse notamment), les statistiques sur les migrations sont produites à l'aide de *systèmes dont le registre de la population constitue une des sources*, mais pas nécessairement la plus importante. En Hongrie, on utilise un système de registres relativement complexe dans lequel les différentes sources de données sont complémentaires. Les données relatives à l'immigration et à l'émigration sont obtenues en intégrant des données provenant de plusieurs sources, notamment de l'Office de l'immigration et de l'asile et du Registre central de la population administré par le Ministère de l'intérieur. L'intégration des données permet de produire des statistiques de grande qualité sur les migrations et de compléter les informations obtenues à partir des sources primaires.

57. En Israël, les données sur les mouvements et les effectifs migratoires proviennent principalement des enregistrements aux frontières. Elles sont communiquées au Bureau central de la statistique par l'Autorité en charge de la population et de l'immigration (qui administre également le registre de la population). Elles sont combinées avec d'autres fichiers administratifs (fichiers fiscaux, fichiers relatifs à l'éducation et autres), avec les données des recensements de la population et avec d'autres données d'enquêtes, afin d'affiner la qualité des estimations produites et d'ajouter des variables provenant d'autres sources et se rapportant à un même individu ou à un même événement.

58. Aux Pays-Bas, les statistiques sur les migrations sont, à l'image de toutes les statistiques sociodémographiques, basées sur le système des bases de données statistiques sociales, fruit de l'intégration des données provenant de plus de 50 registres administratifs. On considère que, à de rares exceptions près, toute la population des Pays-Bas est enregistrée. S'agissant de l'émigration, environ un tiers des personnes qui quittent le pays ne se font pas radier. Lorsqu'elles constatent qu'une personne est manquante, les autorités mènent une enquête et, si la personne demeure introuvable, elle peut être enregistrée comme ayant émigré vers un pays inconnu.

59. En Suisse, les données sur les migrations sont présentes dans trois types de registres principaux : 1) les registres locaux de la population, administrés par les communes ; 2) un registre fédéral dans lequel figurent tous les ressortissants étrangers qui résident légalement dans le pays et sont en possession d'un titre de séjour ; et 3) un registre fédéral spécial pour les diplomates étrangers, les membres des services consulaires, les employés des organisations internationales et les membres de leur famille ainsi que les catégories apparentées. Une forme simplifiée d'intégration des données est réalisée par l'appariement des données recueillies au niveau fédéral et de celles collectées au niveau local. De plus, certaines variables qui ne sont disponibles que dans le registre des résidents étrangers légaux sont ajoutées aux entrées des registres locaux de la population ayant trait aux mêmes observations.

60. En ce qui concerne les *pays ne disposant pas d'un registre de la population*, les principales sources de données sur les migrations sont, notamment, les cartes remplies par les passagers qui se présentent aux frontières, les données figurant dans les passeports, les visas et les permis de séjour. Dans les pays qui disposent d'un registre sur la population, les données provenant de différentes sources sont intégrées de façon à améliorer la qualité des données et l'exhaustivité des informations. L'intégration des données peut, en particulier, contribuer à améliorer les informations relatives à la date et au lieu effectifs d'installation des immigrants, informations qui diffèrent fréquemment de celles qui sont données initialement par les personnes à leur arrivée.

61. En Australie, le Bureau australien de la statistique compile les données relatives aux arrivées et aux départs à partir des informations provenant de différents systèmes de traitement, notamment des cartes remplies par les passagers et des informations figurant

dans les passeports et les visas. Les données sont couplées au moyen d'un code personnel d'identification qui permet aussi au Bureau australien de la statistique de retracer le parcours individuel des voyageurs. Les données sur les migrants sont également couplées avec les données de recensement et les informations fiscales.

62. Au *Canada*, Statistique Canada reçoit des données administratives concernant les immigrants et les résidents non permanents. Étant donné que la province de destination et la date d'installation ne correspondent pas toujours au lieu et à la date d'installation effectifs, l'intégration des données est réalisée à l'aide d'autres sources telles que les fichiers fiscaux. Le couplage des données individuelles s'appuie également sur d'autres sources (recensement de la population du Canada, enquête nationale sur les ménages et statistiques de l'Office des migrations des États-Unis), afin de produire des estimations de qualité et d'obtenir des informations plus complètes sur les caractéristiques socioéconomiques des migrants.

63. En *Nouvelle-Zélande*, les cartes remplies par les passagers sont numérisées et couplées avec les données électroniques sur le franchissement des frontières et les données individuelles des passagers afin d'obtenir des informations supplémentaires et dans un souci de validation. De plus, dans le cas des voyageurs ayant indiqué leur intention de séjourner dans le pays ou de le quitter pour une période prolongée, les données relatives aux voyages et aux passagers sont couplées pour les seize mois précédant le mois de référence, afin de confirmer le statut de résidence en Nouvelle-Zélande à la date du voyage.

64. Au *Royaume-Uni*, il existe un grand nombre de statistiques officielles qui renferment des informations sur les différents aspects des effectifs et des flux migratoires internationaux. Pour ce qui est des flux, la principale source de données est l'enquête internationale sur les passagers (IPS), enquête par sondage dont l'un des multiples objectifs est de collecter des informations auprès des passagers qui entrent dans le pays ou qui le quittent. L'IPS ne prend pas en compte l'évolution des intentions des passagers. C'est pourquoi, pour produire des estimations sur les migrations internationales de longue durée, on procède à l'intégration des données de l'IPS, de l'enquête sur les forces de travail, des données administratives du Ministère de l'intérieur concernant l'immigration, et des estimations des flux à la frontière terrestre de l'Irlande du Nord (lesquels ne sont pas couverts par l'IPS).

65. Aux *États-Unis*, le Census Bureau réalise des estimations annuelles des flux migratoires internationaux en ajoutant divers sous-éléments provenant de sources de données différentes. Une intégration limitée de certains de ces sous-éléments (au niveau des agrégats) est réalisée à l'aide de certaines sources étrangères ou administratives, telles que l'American Community Survey, les recensements étrangers et les mouvements militaires réalisés par le Ministère de la défense. Le Census Bureau ne pratique pas l'intégration des microdonnées sur les migrations internationales, bien que certaines expérimentations aient été réalisées dans ce sens.

66. En résumé, les études de cas montrent que presque tous les pays étudiés ont recours à l'intégration des données pour améliorer la qualité des statistiques sur les migrations et compléter les informations obtenues à partir des sources primaires.

67. Dans les pays disposant d'un registre de la population, l'intégration des données (y compris par l'analyse des « signes de présence ») peut permettre de détecter : 1) les personnes ayant émigré, mais figurant toujours dans le registre de la population ; 2) les personnes ayant immigré, mais ne figurant pas encore dans le registre de la population ; et 3) les personnes décédées.

68. Dans les pays ne disposant pas d'un registre de la population, l'intégration des données peut contribuer à améliorer les informations relatives au lieu et à la date effectifs d'installation des immigrants, lesquelles diffèrent fréquemment de celles qui sont initialement communiquées par les personnes à leur arrivée.

69. Les études de cas ont mis en évidence de nombreux exemples d'intégration des microdonnées et des macrodonnées. Ainsi, dans des pays tels que la Hongrie ou le Canada, on couple des sources de microdonnées différentes. Cette micro-intégration est réalisée en reliant entre elles des sources administratives concernant les étrangers et en reliant ces

mêmes sources au registre de la population afin d'éviter les chevauchements et d'obtenir des variables supplémentaires. D'un autre côté, le Royaume-Uni offre un exemple intéressant d'intégration de macrodonnées : les données de l'enquête réalisée auprès des passagers aux frontières sont combinées, après étalonnage, avec une autre source (l'enquête sur les forces de travail), pour permettre une ventilation territoriale des flux entrants.

70. Comme on pouvait s'y attendre, l'intégration des données est réalisée à partir de nombreux types de sources différents, comprenant divers registres administratifs ou statistiques, d'autres données administratives (données issues des contrôles aux frontières) et des enquêtes statistiques (recensements). Peu de pays ont recours aux statistiques miroirs des pays de destination pour améliorer leurs estimations des flux d'émigration. Cette démarche concerne principalement les agrégats, même s'il existe quelques exemples d'échanges internationaux de microdonnées, notamment entre les pays nordiques (lesquels ne sont pas couverts par les études de cas) ou limités aux pays voisins.

V. Métadonnées

71. Lorsque l'on compare les estimations des migrations réalisées par différents pays, il est très important de disposer de métadonnées complètes afin d'évaluer en détail la comparabilité des données.

72. Théoriquement, les métadonnées peuvent être scindées en trois parties : la première, propre à chaque source de données, donne une description des principales caractéristiques de ces sources et de la manière dont les ensembles de données sont transformés avant d'être intégrés ; la deuxième indique de quelle manière les ensembles de données provenant des sources de données répertoriées sont intégrés ; la troisième se fonde sur quelques mesures de la qualité, essentiellement dans le but de prouver que le contenu informatif s'est étoffé grâce à l'intégration. Ces informations devraient être fournies pour toutes les statistiques migratoires, à savoir celles sur les effectifs de migrants, les flux migratoires et tous leurs sous-groupes présentant un intérêt.

73. La première information à fournir est la liste des sources de données utilisées pour produire les données intégrées. Dans cette liste, les métadonnées suivantes pourraient accompagner chaque source de données :

- a) Nom ;
- b) Typologie (registre administratif, enquête par sondage, recensement, mégadonnées, etc.) ;
- c) Propriétaire (y compris l'autorité statistique d'un pays tiers) ;
- d) Règles d'accès/de fourniture de données (si le propriétaire n'est pas l'institut national de la statistique) ;
- e) Population de référence ;
- f) Date/période de référence des données ;
- g) Fréquence de mise à jour ;
- h) Degré d'actualité (soit le délai entre la date de disponibilité des données et la date/période de référence) ;
- i) Niveau de détail des données originales (microdonnées, macrodonnées) ;
- j) Niveau de détail de l'ensemble de données utilisé pour l'intégration (microdonnées, macrodonnées) ;
- k) Description de toutes les méthodes employées pour transformer l'ensemble de données d'origine en données d'entrée susceptibles d'être intégrées ;
- l) Dimensions de l'ensemble de données (n x p) ;
- m) Variables contenues dans la source de données (ou large description, si elles sont nombreuses) ;

n) Variables contenues dans les sources de données qui sont conservées pour être intégrées (éventuellement sous-ensemble de l'élément précédent) ;

o) Variables utilisées pour intégrer cette source de données aux autres (éventuellement sous-ensemble de l'élément précédent) ;

p) Disponibilité du numéro d'identification personnel (PIN) (en principe, cette métadonnée figure dans l'élément précédent mais, compte tenu de son importance, il peut être utile de lui accorder une place particulière) et de ses principales caractéristiques, le cas échéant.

74. À supposer que le processus d'intégration soit mis en œuvre dans l'ordre et par étapes, c'est-à-dire à mesure que chaque paire d'ensembles de données est intégrée, les informations suivantes pourraient être fournies :

a) Numéro d'étape ;

b) Ensembles de données concernés (nom ou numéro figurant sur la liste des sources de données individuelles) ;

c) Périodicité de la procédure d'intégration ;

d) Variables employées pour l'intégration (le cas échéant) ;

e) Description de la méthode d'intégration des ensembles de données ;

f) Chevauchements de données ;

g) Principaux problèmes ou difficultés rencontrés lors de l'intégration de ces ensembles de données, tels que l'établissement de rapports sur l'absence de couplage, ainsi que les méthodes pour évaluer les taux de rapprochement faux positifs et les résultats de cette évaluation ;

h) Dimensions (n x p) de l'ensemble de données intégrées obtenu.

75. Dans certains cas, notamment lorsque les statistiques migratoires sont établies à partir des registres de la population, le nombre de sources de données peut être relativement élevé, ce qui peut rendre contraignante la communication des métadonnées. Dans de tels cas, il peut être utile d'établir un seul rapport général sur le fonctionnement du système. Toutefois, il conviendrait également de tenir compte du fait qu'il peut être important pour les utilisateurs de comprendre clairement comment les données définitives sont produites au moyen d'un système de gestion des migrations qui peut être complexe. La complexité ne doit pas empêcher la transparence.

76. La troisième catégorie de métadonnées, qui exige un travail beaucoup plus poussé, peut s'articuler autour de la première série de mesures quantitatives de la qualité énoncées ci-dessous, afin de mettre en évidence l'effet positif de l'intégration des données :

a) Différence au niveau du nombre des données enregistrées ou de la taille de la population visée entre l'ensemble final de données intégrées et le plus petit ensemble de données individuel : $n^* - \min(n)$;

b) Différence au niveau du nombre de variables entre l'ensemble final de données intégrées et le plus petit ensemble de données individuel : $p^* - \min(p)$;

c) Différence au niveau du degré d'actualité entre l'ensemble final de données intégrées et l'ensemble de données individuel le moins actuel : $t^* - \max(t)$;

d) Différence au niveau du nombre de données enregistrées ou de la taille de la population visée entre l'ensemble final de données intégrées et le plus grand ensemble de données individuel : $n^* - \max(n)$;

e) Différence au niveau du nombre de variables entre l'ensemble final de données intégrées et le plus grand ensemble de données individuel : $p^* - \max(p)$;

f) Différence au niveau du degré d'actualité entre l'ensemble final de données intégrées et l'ensemble de données individuel le plus actuel : $t^* - \min(t)$;

g) Gain dans la qualité des estimations, tel que le pourcentage de diminution de la variance des estimations ou toute mesure quantitative d'une amélioration de la précision.

77. Les métadonnées peuvent être enrichies par une évaluation qualitative des effets de l'intégration, telles que des observations concernant l'amélioration des sources de données initiales et l'allègement de la charge que constitue la communication de données, qui peut être étayée, le cas échéant, par des références/exemples spécifiques et par une conclusion générale sur le processus d'intégration.

VI. Conclusion

78. Les statistiques migratoires constituent probablement la matière la plus compliquée dans le domaine des statistiques sociales, non seulement sous l'angle opérationnel, mais aussi d'un point de vue théorique. S'il est vrai qu'il existe des définitions internationales, il n'est pas toujours aisé de préciser la notion de migrant au sens strict. En outre, la migration prend des formes toujours plus variées, et il existe diverses sources pour effectuer une mesure partielle des migrations.

79. La difficulté pour les organismes de statistique est d'associer des données provenant de différentes sources à des informations sur les migrations qui soient fiables et intégrées. Les 13 études de cas montrent les différents moyens et sources de données que les pays utilisent pour améliorer leurs statistiques migratoires.

80. L'Équipe spéciale a mené une enquête pour évaluer la capacité de nombreuses sources différentes à produire des statistiques migratoires. D'après les réponses reçues, la plupart des pays (31 sur les 56 qui ont répondu) ont recours à *plus d'une source pour établir des statistiques migratoires*. Les différences d'approche s'expliquent par la grande diversité des situations et des sources de données d'un pays à l'autre. La tendance générale consistant à faire appel à davantage de sources administratives pour mesurer les migrations se manifeste tant dans les pays qui emploient les méthodes traditionnelles (enquêtes, recensements exhaustifs) que dans ceux qui se fondent sur les registres de la population pour établir leurs statistiques.

81. Le choix des sources de données sur les migrations dépend souvent de la *géographie*. Les pays relativement isolés ou qui ont moins de points d'entrée et de sortie (par exemple, la Nouvelle-Zélande et Malte) peuvent établir des statistiques migratoires au moyen de sources très différentes, contrairement aux pays où les points d'entrée sont nombreux ou les postes de contrôle inexistant, comme les pays de l'espace Schengen en Europe.

82. L'existence de *différentes définitions* données par les sources administratives au terme « population migrante » constitue une autre limite importante pour l'étude des statistiques migratoires. Dans de nombreux cas, le délai normal de douze mois de séjour effectif ou prévu n'est pas celui qui est utilisé pour établir les statistiques migratoires car d'autres conceptions (par exemple d'autres délais) ont été adoptées. En outre, il se peut que différentes sources administratives à l'intérieur d'un même pays communiquent des informations fondées sur des concepts différents.

83. Dans les pays de l'Union européenne, les statistiques migratoires doivent être *pleinement compatibles avec l'évolution des effectifs de migrants*. Dans de nombreux autres cas, les statistiques migratoires n'ont pas pour objectif cette pleine adéquation entre les flux et les effectifs. De plus, de nombreux pays produisent des données distinctes sur les migrants de longue durée, les demandeurs d'asile, la population étrangère non résidente, les migrants de courte durée, etc.

84. S'agissant de l'intégration des données, il existe quelques exemples intéressants de combinaisons *tant au niveau des macrodonnées qu'à celui des microdonnées*. En ce qui concerne la micro-intégration, à savoir la combinaison de sources au niveau d'un enregistrement individuel, plusieurs exemples peuvent illustrer la tendance pour l'avenir. La combinaison des données enregistrées peut engendrer des chevauchements qui sont pris en compte dans le cadre des processus d'ajustement et d'étalonnage. Cette intégration vise à mieux mesurer les effectifs mais aussi à établir des statistiques plus complètes sur les flux.

85. La principale faiblesse des registres de la population est la *difficulté à repérer l'émigration*. Si une personne n'informe pas les autorités qu'elle a quitté un pays, les registres administratifs doivent être mis à jour dans le cadre de procédures administratives qui ne sont pas toujours simples ou immédiates, si bien que des personnes qui ne résident plus dans le pays peuvent continuer de figurer dans les registres de la population.

86. En combinant les sources, on peut détecter des « *signes de présence* ». Ainsi, on peut croiser un registre de la population avec des fichiers fiscaux ou de sécurité sociale afin d'avoir davantage l'assurance que certaines personnes résident dans le pays. Cela revêt une importance particulière lorsque différents enregistrements administratifs sont combinés, comme c'est le cas en Lettonie ou en Autriche, où l'on recourt à des régressions logistiques pour évaluer le nombre de résidents à partir d'éléments qui figurent dans différents registres administratifs. L'analyse des « *signes de présence* » peut aussi permettre d'identifier des personnes qui ont émigré mais dont le nom ne figure pas dans le registre de la population (Italie) ainsi que des personnes décédées (Autriche et Espagne).

87. Pour ce qui est des statistiques sur les flux migratoires, il existe quelques exemples d'intégration au sens strict, c'est-à-dire de combinaison de sources visant à obtenir une base de données unique. Par exemple, en Hongrie et au Canada, différentes sources sont couplées au niveau des microdonnées. On procède à cette micro-intégration en couplant les informations sur les étrangers provenant de différentes sources administratives avec celles du registre de la population, afin d'éviter les chevauchements et de fournir d'autres variables.

88. Une initiative que les utilisateurs commencent à solliciter et qui pourrait être très prometteuse est l'*échange de données individuelles* entre les pays à des fins statistiques. La principale conclusion à tirer de la mesure des flux migratoires est que ces déplacements produisent des effets dans le pays d'origine et le pays de destination. Il est donc logique de penser que le partage des données statistiques permettra d'améliorer ces statistiques. Les échanges de données individuelles à des fins statistiques sont susceptibles de se multiplier, même s'il est nécessaire de surmonter des obstacles qui ne sont pas seulement juridiques ou techniques. Les pays nordiques fournissent un bon exemple à cet égard.

VII. Recommandations

89. Sur la base des travaux menés par l'Équipe spéciale et du contenu du présent document, un certain nombre de recommandations adressées aux organismes nationaux de statistique sont présentées dans le présent chapitre. Ces recommandations sont jugées utiles pour tous les pays, indépendamment de leur situation actuelle en ce qui concerne la disponibilité et l'utilisation de données administratives pour l'élaboration de statistiques migratoires.

A. Améliorer l'accès aux données administratives pour les organismes nationaux de statistique

90. Dans les pays où les registres administratifs ne sont pas consultables ou ne le sont que dans une certaine mesure par les organismes de statistique qui établissent des statistiques officielles, l'une des priorités consiste à prendre des mesures pour faciliter l'accès aux données administratives et pour généraliser leur utilisation par les organismes de statistique. Un appui politique et, dans certains cas, des modifications de la législation pourraient donc être nécessaires.

B. Utiliser les données administratives pour établir des statistiques migratoires

91. De nombreuses sources administratives contiennent des informations susceptibles d'être utilisées dans les statistiques migratoires. Il s'agit notamment des registres et des bases de données sur la population, les visas, les passeports, les contrôles aux frontières, les

autorisations de séjour et de travail ainsi que les demandeurs d'asile et les réfugiés. Il est recommandé d'examiner le potentiel de toutes ces sources administratives et de l'exploiter pour établir des statistiques migratoires, pour autant que ces sources soient utilisables à cette fin. Cela exige une coopération avec les ministères, les organismes et les administrations chargées de recueillir des données.

92. Il est souhaitable d'élaborer plusieurs produits statistiques qui donnent des informations sur les effectifs de migrants, ainsi que sur les flux d'entrée et de sortie, afin de viser différentes catégories de flux migratoires. Dans certains cas, des statistiques différentes peuvent fournir des données qui ne sont pas pleinement cohérentes. Cependant, leur publication peut contribuer à brosser un tableau général de la situation et à faire prendre conscience qu'un produit unique ne donne pas toutes les informations nécessaires sur les migrations internationales. La simple publication de données provenant de sources diverses est un premier moyen d'intégration des informations sur les migrants.

C. Combiner les données provenant de différentes sources au moyen des « signes de présence »

93. Dans les pays où les données administratives peuvent être exploitées pour établir des statistiques, la combinaison de sources utilisant des « signes de présence » est de très bon augure. Le point commun à ces différentes approches est l'utilisation de plusieurs sources faisant état de la présence d'une personne dans un pays pour évaluer la probabilité que celle-ci y réside. Une telle combinaison de données provenant de différentes sources permet d'améliorer les estimations du nombre d'habitants à un moment donné et du nombre d'étrangers.

D. Veiller à la qualité des données intégrées

94. Il est probable que l'intégration de données provenant de sources multiples au niveau des unités individuelles tendra à se renforcer au fur et à mesure que les pays progresseront dans la normalisation des registres de statistiques sociodémographiques. L'évaluation de la qualité des données intégrées est appelée à faire l'objet d'une attention accrue dans le contexte de la production de statistiques relatives aux migrations internationales.

E. Présenter de manière transparente les méthodes d'intégration des données et élaborer des normes

95. Il est de plus en plus important que les organismes de statistique présentent de manière transparente les règles et les méthodes qu'ils appliquent pour mener à bien les processus d'intégration de données qu'ils mettent en œuvre pour produire des statistiques relatives aux mouvements migratoires internationaux et d'autres formes de statistiques concernant les migrations internationales. Cela étendra en effet la possibilité d'élaborer de nouvelles normes méthodologiques d'intégration de données applicables par les différents organismes nationaux de statistique. Chaque pays disposera toujours de ses propres pratiques en matière de collecte et d'intégration des données, mais il est probable que davantage de lieux de rencontre ouverts se créeront pour examiner les règles et méthodes employées pour combiner les sources utilisées pour produire des statistiques sur les migrations.

F. Favoriser la comparaison internationale et l'échange de données sur les migrations

96. La comparaison internationale et l'échange de données sur les migrations, les personnes nées à l'étranger ou les étrangers peuvent constituer une autre source d'amélioration. En général, les sources administratives sous-estiment l'émigration en raison de la faible motivation des émigrants pour faire enregistrer leur sortie du pays. Toutefois,

dans les pays de destination, les données sur l'immigration pourraient rendre compte de ces déplacements. Cela peut être vérifié grâce à la méthode dite des « statistiques miroirs », qui rend compte des flux migratoires du pays A vers le pays B sur la base des statistiques sur l'émigration produites par le pays A et des statistiques sur l'immigration produites par le pays B. En outre, les données des pays de destination peuvent fournir des informations sur le nombre et les caractéristiques des émigrants et des ressortissants d'un pays d'origine. L'échange de données avec d'autres pays peut donc s'avérer très utile pour comprendre les processus migratoires et améliorer les systèmes de mesure des migrations.

97. Les sources internationales de données présentent une très grande utilité à cet égard. En ce qui concerne l'Europe, Eurostat publie des données très détaillées sur les flux d'entrée et de sortie qui permettent d'effectuer cette comparaison entre les pays de l'Union européenne. Dans le cadre de son Mécanisme d'échange d'informations sur les statistiques relatives aux migrations (base de données), la CEE publie ces données provenant des pays d'Europe orientale, du Caucase et d'Asie centrale.

98. Des résultats décisifs pourraient être obtenus si les bases de données étaient partagées entre les pays pour établir des corrélations à l'échelle individuelle. Cette pratique a déjà été mise en œuvre dans les pays nordiques. Aujourd'hui, il est peu probable que les échanges de données au niveau individuel se généralisent, mais il se pourrait qu'en raison de l'intensité de leurs échanges ou d'autres considérations, certains pays prennent des initiatives dans ce sens. La réussite de ces échanges pourrait encourager d'autres pays à suivre la même voie.

VIII. Travaux futurs

99. Au niveau international, d'autres travaux pourraient être menés sur la possibilité d'utiliser les mégadonnées pour établir des statistiques migratoires ainsi que sur la manière dont l'analyse des données longitudinales peut être intégrée dans les statistiques migratoires annuelles et peut compléter les mesures transversales disponibles.

A. Le potentiel des mégadonnées

100. Les sources de données nouvelles et non conventionnelles, comme les téléphones mobiles, les cartes de crédit et les réseaux sociaux – communément appelées mégadonnées – pourraient s'avérer utiles pour établir des statistiques migratoires. À l'heure actuelle, les organismes de statistique ont de grandes difficultés à accéder à ces données, qui sont le plus souvent détenues par des entreprises privées. Dans les années à venir, des mesures devraient être prises pour accéder à ces sources et trouver les moyens de les utiliser aux fins des statistiques officielles. À elles seules, ces informations seraient probablement insuffisantes et partiales, mais une fois intégrées dans des sources traditionnelles, elles pourraient améliorer les estimations des migrations. Il serait important de mettre en commun ces nouvelles pratiques au niveau international, afin de soutenir les premières mesures prises par certains pays pour exploiter le potentiel de ces données dans le domaine des statistiques migratoires.

B. Données longitudinales

101. Il est probable qu'à l'avenir l'intégration des données ne se résumera pas à améliorer la qualité de la mesure des flux migratoires. L'intégration des données permettra peut-être d'accéder à de nouvelles données sur les migrants. La possibilité de relier les bases de données (ou les enquêtes) démographiques avec des fichiers concernant l'emploi, la sécurité sociale, les impôts, la santé, les systèmes d'enseignement, et bien d'autres domaines encore, permettra d'obtenir à l'avenir des informations plus nombreuses et de meilleure qualité sur les conditions de vie ou l'intégration sociale des migrants.

102. Les informations figurant dans les études longitudinales en sont un exemple éloquent. Les statistiques officielles mettent l'accent presque exclusivement sur la fourniture d'informations transversales sur les migrants. Par exemple, il n'existe guère

d'informations sur la durée des épisodes migratoires (la question de savoir si un migrant vit dans le pays d'accueil depuis longtemps), ou sur les périodes qui revêtent une très grande importance lorsque l'on étudie les flux migratoires, à savoir le temps écoulé entre l'arrivée dans le pays de destination et l'obtention d'un emploi ou entre la perte d'un emploi et l'émigration. Les données longitudinales sont nécessaires pour répondre à la question de savoir si la situation socioéconomique s'améliore ou s'aggrave avec le temps. Elles sont également nécessaires pour comprendre la migration familiale et l'émergence de schémas tels que la migration circulaire.

103. La mise en commun des bonnes pratiques pour créer des ensembles de données dans une perspective longitudinale et leur application dans les statistiques sur la migration régulière aideraient les instituts nationaux de la statistique dans leurs travaux. L'élaboration de recommandations sur la manière d'harmoniser ces statistiques au niveau international pourrait être l'étape suivante.

Références

- ESSnet (2008) : Projet relatif à l'intégration des données d'enquête et des données administratives. Documentation consultable à l'adresse https://ec.europa.eu/eurostat/cros/content/isad-0_en.
- ESSnet (2011) : Projet sur l'intégration des données. Documentation consultable à l'adresse https://ec.europa.eu/eurostat/cros/content/data-integration_en.
- ESSnet (2014) : Projet relatif au document intitulé *Handbook on Methodology of Modern Business Statistics (Memobust handbook), Module on Macro Integration*. Documentation consultable à l'adresse https://ec.europa.eu/eurostat/cros/content/macro-integration_en.
- Eurostat (2009) : *Insights on Data Integration Methodologies. Proceedings of the ESSnet-ISAD workshop, Vienna, 29–30 May 2008. Eurostat Methodologies and Working Papers*. Disponible à l'adresse : <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-09-005>.
- Eurostat (2013) : *Statistical matching : a model based approach for data integration. Eurostat Methodologies and Working Papers*. Disponible à l'adresse : <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-13-020>.
- SDMX (2009) : *SDMX Content-Oriented Guidelines – Annex 4 : Metadata Common Vocabulary*. Disponible à l'adresse <http://www.sdmx.org>.
- Groupe de haut niveau de la CEE sur la modernisation de la statistique officielle (2017) : Examen approfondi de l'intégration des données. Document ECE/CES/2017/8 rédigé pour la réunion de la Conférence des statisticiens européens, 19-21 juin 2017.
-