

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

Genève, 22-24 juin 2020

Point 4 f) de l'ordre du jour provisoire

Rapports, directives et recommandations élaborés**sous les auspices de la Conférence :****Données longitudinales sur les migrations****Indications relatives à l'utilisation de données longitudinales
pour les statistiques migratoires¹****Note de l'Équipe spéciale de l'utilisation de données longitudinales
pour les statistiques migratoires***Résumé*

Les présentes indications ont été élaborées par l'Équipe spéciale de l'utilisation de données longitudinales pour les statistiques migratoires, composée de représentants du Canada (présidence), de l'Allemagne, de l'Autriche, de la Belgique, de l'Espagne, de la Fédération de Russie, de l'Italie, du Kazakhstan, du Mexique, des Pays-Bas, du Royaume-Uni, de la Suisse, de la Turquie, d'Eurostat, de l'OCDE et de la CEE.

La version abrégée des indications a été établie à des fins de traduction. Elle comprend une introduction, les sections introductives de tous les chapitres de fond ainsi que l'intégralité du chapitre comportant les conclusions, les recommandations et les travaux complémentaires à mener.

Le texte intégral des indications a été adressé par voie électronique à tous les membres de la Conférence des statisticiens européens pour consultation. Il est disponible sur la page Web de la Conférence : <http://www.unece.org/index.php?id=53381>.

Si la consultation donne des résultats positifs, la Conférence sera invitée à approuver les indications à sa réunion plénière.

¹ Il a été convenu que le présent document serait publié après la date normale de publication en raison de circonstances indépendantes de la volonté du soumetteur.



I. Chapitre 1 : Introduction

1. Le nombre de migrants internationaux ne cessant de croître, il est de plus en plus important pour le public et les décideurs de comprendre les flux migratoires et l'incidence des migrations sur les individus, les familles, les sociétés et les économies. Dans de nombreux cas, les questions clés concernent le processus d'établissement des migrants – durée de leur séjour dans le pays d'accueil, manière dont ils s'intègrent dans les sociétés d'accueil et évolution de leur situation socioéconomique dans le temps.
2. En fin de compte, étudier les migrations revient à étudier les changements. Les changements concernent d'abord la résidence, mais peuvent s'étendre au statut juridique ou au statut de résident ainsi qu'à la situation socioéconomique. La migration, l'intégration et l'établissement sont des processus, pas des états, et les résultats peuvent apparaître à court ou à long terme. En raison de ce caractère transitoire, ce sont des thèmes qui se prêtent bien à une étude longitudinale. Il a été souligné dans le Pacte mondial pour des migrations sûres, ordonnées et régulières qu'une base temporelle est indispensable pour les statistiques migratoires et que des données sont nécessaires pour « suivre et évaluer efficacement le respect des engagements au fil du temps ».
3. Les méthodes traditionnelles de collecte de données longitudinales (par exemple, les études de cohortes et les enquêtes par panel) sont devenues plus difficiles à appliquer en raison des coûts élevés et du niveau important d'attrition. Par conséquent, les pays se tournent de plus en plus vers d'autres sources de données pour produire des résultats longitudinaux.
4. La production de statistiques a évolué dans de nombreux pays du fait que les données administratives sont plus disponibles et faciles à utiliser. Les banques de données administratives ne cessent de s'améliorer en termes d'exhaustivité, de fréquence et de qualité des informations recueillies. De plus en plus de pays peuvent constituer des ensembles de données longitudinales sans supporter des coûts excessifs en généralisant les données administratives et l'intégration des données pour produire des statistiques migratoires.
5. Alors que par le passé, les études longitudinales sur la migration étaient souvent des études indépendantes ou ponctuelles, les organismes nationaux de statistique commencent maintenant à les intégrer de plus en plus dans leur production régulière de statistiques migratoires. Toutefois, il n'existe actuellement aucune directive internationale expliquant comment créer des sources de données longitudinales à partir de données intégrées, relever les divers défis posés par de tels projets et diffuser les indicateurs clés et autres résultats.
6. Le recours à une méthode longitudinale pour mesurer l'intégration des migrants a été discuté à plusieurs reprises à la réunion de travail CEE-Eurostat sur les statistiques migratoires (par exemple en 2012, 2014 et 2017). En octobre 2017, il a été recommandé à la réunion de travail de poursuivre les travaux méthodologiques à ce sujet, d'examiner les pratiques nationales et d'élaborer des recommandations favorisant la comparabilité internationale des données longitudinales.
7. En février 2018, le Bureau de la Conférence des statisticiens européens a créé, dans le cadre de la CEE, une équipe spéciale de l'utilisation de données longitudinales pour les statistiques migratoires, dont l'objectif était de préparer des directives sur la manière d'intégrer des données longitudinales dans les statistiques migratoires annuelles et de compléter les mesures transversales disponibles. Le présent document reproduit les résultats de l'Équipe spéciale.
8. Les présentes indications s'appuient sur les récents travaux méthodologiques menés par des équipes spéciales de la CEE. La publication « Measuring change in the socio-economic characteristics of migrants » (Mesurer l'évolution des caractéristiques socioéconomiques des migrants) (CEE, 2015) montrait en quoi il était avantageux d'utiliser des données longitudinales et recommandait aux pays de développer des méthodes de couplage de données pour acquérir des ensembles de données longitudinales. La publication ultérieure « Guidance on data integration for measuring migration » (Document d'orientation sur l'intégration des données pour mesurer les migrations) (CEE, 2019 b)) présentait plusieurs exemples où l'intégration de différents ensembles de données avait conduit à la compilation de données longitudinales.

9. Les présentes indications se composent de trois principaux chapitres de fond :
- Chapitre 2 : Aperçu des sources de données longitudinales destinées aux statistiques migratoires
 - Chapitre 3 : Comment élaborer un ensemble de données longitudinales pour les statistiques migratoires au moyen de données intégrées
 - Chapitre 4 : Diffusion de statistiques migratoires régulières à partir de sources de données longitudinales
10. Les chapitres constituent des parties autonomes et il n'est pas nécessaire de les lire conjointement avec les autres parties du rapport. L'important est que les utilisateurs tirent parti des parties les plus pertinentes pour leur travail.
11. Le chapitre 2 présente un aperçu des sources de données longitudinales destinées aux statistiques migratoires. Il donne des instructions sur ces différents types de données ainsi que des exemples concrets émanant de divers pays. Une section spéciale dispense des conseils sur l'utilisation des registres de la population aux fins des statistiques migratoires, notamment concernant les difficultés particulières et les moyens d'en venir à bout.
12. Le chapitre 3 présente un guide montrant comment élaborer des ensembles de données longitudinales pour les statistiques migratoires au moyen de données intégrées. S'agissant des travaux menés par des équipes spéciales précédentes, il indique une voie à suivre depuis la conception des statistiques jusqu'à leur diffusion, traitant notamment des problèmes souvent rencontrés et donnant des exemples concrets de pratiques optimales suivies par divers pays. Les phases visées comprennent les activités suivantes :
- Conception statistique ;
 - Évaluation et prétraitement des fichiers sources ;
 - Intégration de données destinées aux données longitudinales ;
 - Attribution d'identificateurs longitudinaux individuels ;
 - Création de la base de données finale ;
 - Diffusion des résultats.
13. Le chapitre 4 propose un ensemble d'indicateurs longitudinaux relatifs aux migrations et présente les meilleures pratiques à suivre pour la diffusion des résultats longitudinaux. Chaque indicateur comprend un résumé des difficultés et des meilleures pratiques, assorti d'exemples émanant de divers pays.
14. L'un des thèmes principaux des présentes indications est le fait que l'élaboration et l'utilisation de données longitudinales pour les statistiques migratoires sont complexes et difficiles. Les différents chapitres montrent cependant comment on pourrait remédier à ces limites. Le chapitre 4 souligne comment il est encore possible de produire des indicateurs de valeur malgré les limites de la base de données obtenue. Afin de répondre aux besoins émergents et croissants pour ce qui est de mieux comprendre les caractéristiques de la migration, de l'intégration et de l'établissement des migrants internationaux, il est important que les organismes nationaux de statistique comprennent les sources de données disponibles, examinent comment leurs limites peuvent être atténuées ou traitées et proposent des méthodes de diffusion des résultats qui tiennent compte de ces caractéristiques.
15. Une plus grande utilisation des données intégrées et des registres statistiques permettra à de nombreux organismes nationaux de statistique de bénéficier d'un immense potentiel inexploité s'agissant des statistiques longitudinales sur les migrations. Le chapitre 3 indique la solution qui permettra aux pays de créer de nouvelles sources de données longitudinales. Le chapitre 4 donne des exemples de ce qui peut être fait avec les résultats obtenus.
16. L'étude des migrations est finalement une application naturelle de l'analyse longitudinale. L'intensification du recours à l'intégration des données et aux registres statistiques ouvrent de nouvelles perspectives en matière d'élaboration et de diffusion de données longitudinales pour les statistiques migratoires. Certes, il existe des difficultés, mais

les avantages que présente l'exploitation de ces nouvelles possibilités sont bien plus importants. Étudier les migrations revient à étudier les changements et les données longitudinales sont essentielles à cette fin.

II. Chapitre 2 : Aperçu des sources de données longitudinales destinées aux statistiques migratoires

A. Définition

17. Les données longitudinales s'entendent des informations qui sont collectées dans le temps auprès des mêmes unités d'analyse, telles que les individus ou les ménages. L'analyse longitudinale permet de décrire de manière unique et précise les trajectoires individuelles dans le temps. Les données longitudinales permettent d'étudier et de comprendre les événements et les transitions qui se produisent au cours de la vie et d'une génération à l'autre. Cela est particulièrement utile pour l'étude des migrations internationales, car l'établissement dans un nouveau pays constitue un processus à long terme. Les données longitudinales peuvent révéler les résultats géographiques et socioéconomiques des migrations. Les données longitudinales collectées sur une longue période permettent de comprendre non seulement les expériences des migrants mais aussi celles de leurs enfants.

18. Le présent chapitre décrit différents types de données longitudinales permettant d'étudier les migrations internationales. On y examine aussi les données transversales pour déterminer comment les utiliser afin d'offrir une perspective longitudinale sur les migrations et les expériences des groupes de migrants. La complémentarité et la comparabilité des mesures longitudinales et transversales sont également étudiées, ainsi que les modèles qui associent les deux approches.

19. On pourrait citer comme exemples les modèles de panel intégrés dans des enquêtes transversales ou encore les questions rétrospectives visant à obtenir des preuves longitudinales dans une enquête transversale. Les cohortes synthétiques des recensements ou enquêtes successifs reproduisent d'une certaine manière les études longitudinales, mais ne suivent pas les mêmes cas dans le temps.

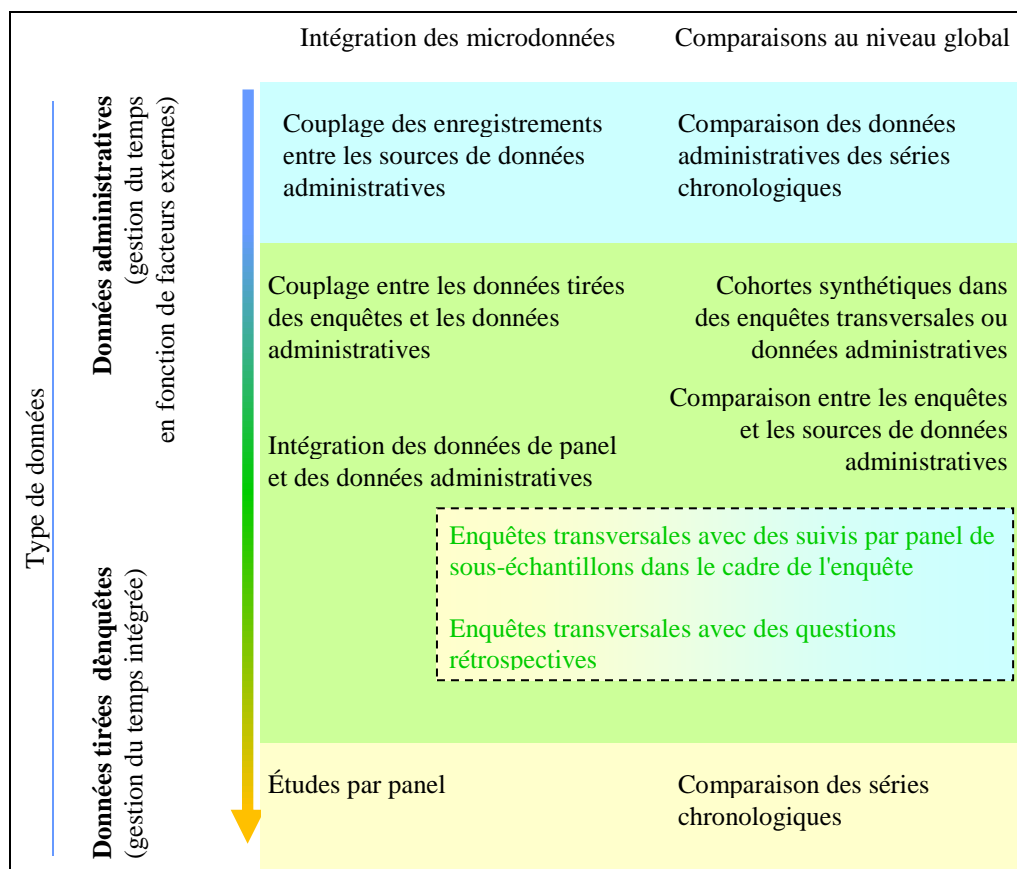
20. Enfin, comme de nombreux pays s'appuient sur les registres de la population pour calculer les chiffres migratoires, les caractéristiques de ces registres qui font qu'ils sont plus ou moins adaptés à l'évaluation des aspects longitudinaux des migrations sont également étudiées de manière assez détaillée.

B. Types de fichiers de données

21. On trouvera dans le schéma ci-dessous une typologie conceptuelle des informations longitudinales relatives aux statistiques sur les migrations internationales.

Figure 1

Typologie des données longitudinales et des perspectives concernant les statistiques migratoires



22. Comme le montre la figure 1, les données peuvent être classées selon qu'elles couplent des informations concernant les mêmes unités dans le temps, et sont donc des données longitudinales, ou selon qu'elles offrent une certaine perspective longitudinale. L'« intégration des microdonnées » a trait aux données qui sont couplées au niveau de l'enregistrement, dans le temps. Les « comparaisons au niveau global » concernent les *perspectives* longitudinales que l'on obtient en comparant des séries chronologiques ou se servant de cohortes synthétiques. Contrairement à l'intégration des microdonnées, les comparaisons au niveau global n'exigent pas que les mêmes individus soient échantillonnés pour chaque vague de collecte de données.

23. Les méthodes appliquées s'inscrivent dans un ensemble homogène avec un équilibre variable entre l'utilisation de données tirées d'enquêtes (en jaune dans le schéma) et de données administratives (en bleu). Celles qui recourent à la fois à des données tirées d'enquêtes et à des données administratives apparaissent en vert ; elles concernent notamment les enquêtes intégrées et les données administratives, ainsi que des cas hybrides enquêtes/données administratives). Avec les données tirées d'enquêtes, les organismes nationaux de statistique peuvent généralement contrôler non seulement les données qui sont collectées, mais aussi la chronologie de la collecte (dans la limite des ressources attribuées). Avec les sources administratives, ces organismes exercent un contrôle moindre sur la collecte des données et doivent aligner au mieux le référencement temporel afin d'atteindre les objectifs de recherche ou d'analyse souhaités.

24. Plus loin dans le présent chapitre, on examinera deux dimensions du temps auxquelles il faudra bien réfléchir lorsque les données administratives seront utilisées longitudinalement. La première est l'opérationnalisation *endogène* du temps dans le modèle de recherche, ce qui signifie que les observations sont faites au moment où les mesures ont effectivement lieu. La seconde est la manifestation *exogène* du temps dans l'ensemble de données administratives, ce qui suppose que les enregistrements sont effectués conformément aux pratiques administratives, à la réglementation, à l'infrastructure et à la qualité de l'administration. Les

disparités entre l'opérationnalisation endogène et la manifestation exogène du temps doivent être comprises et sont une source de biais potentiels dans les données. La question de la radiation en est un exemple explicite. Il en est ainsi lorsque l'émigration n'est pas enregistrée dans les données administratives, et de nombreuses sources, en particulier les registres de la population, connaissent ce problème.

25. Les nouvelles études prospectives par panel supposent le choix d'un échantillon (d'individus ou de ménages, mais aussi par exemple d'entreprises ou d'événements). L'échantillon fait l'objet d'une enquête puis d'un suivi à intervalles réguliers. Les données longitudinales des enquêtes par panel pourraient également prendre la forme de nouvelles données rétrospectives. Dans ce type de données, les répondants sont interrogés notamment à propos d'événements qui se sont produits dans le passé. Un modèle prospectif assorti d'un repère rétrospectif combine les deux concepts, une étude de référence rétrospective étant menée au début puis s'accompagnant d'un suivi prospectif dans le temps.

26. Les organismes nationaux de statistique font de plus en plus usage de sources de données administratives pour l'analyse statistique. L'analyse de la situation des migrants pourrait se fonder entièrement sur des dossiers administratifs, grâce à un couplage des enregistrements visant à créer des ensembles de données longitudinales et à comparer des séries chronologiques de données administratives.

27. Une approche mixte peut prendre plusieurs formes, par exemple en partant des informations provenant d'une enquête et en les actualisant au moyen de dossiers administratifs. Le tableau 1 ci-dessous décrit la création de données longitudinales à partir du couplage des données issues d'un recensement et des dossiers administratifs (études longitudinales faites au Royaume-Uni). On peut aussi associer des modèles transversaux et longitudinaux, comme cela a été le cas dans l'enquête de l'Union européenne sur la population active. Les questions rétrospectives posées dans des enquêtes transversales donnent aussi un aperçu des processus longitudinaux.

28. Si des informations longitudinales tirées de mesures répétées des mêmes individus font défaut, on peut obtenir des perspectives longitudinales à partir de cohortes synthétiques. Les cohortes peuvent être définies comme des « groupes de personnes marchant ensemble dans le temps » ; ces personnes peuvent être définies par leur naissance ou la période de leur arrivée dans un pays, par exemple.

Pratiques actuelles

29. On trouvera ci-après un schéma classant les pratiques internationales actuelles en matière de collecte de données longitudinales sur les migrants.

Figure 2

Typologie des pratiques internationales actuelles fondées sur des données longitudinales et perspectives concernant les statistiques migratoires

Type de données	Intégration des microdonnées	Comparaisons au niveau global		
Données administratives (gestion du temps en fonction de facteurs externes)	Registre central de la population belge	Comparaisons canadiennes des salaires d'entrée dans le temps fondées sur la BDIM		
	Corrélation de données administratives d'Immigration, Réfugiés et Citoyenneté Canada (IRCC)			
	Banque de données longitudinales canadienne sur les immigrants (BDIM)			
	Registre central allemand des étrangers			
	Registres de la population italiens (ANAGRAFE)			
	Permis de séjour en Italie			
	Registre de la population espagnol (Padrón)			
	Base de données longitudinales suisse			
	Données tirées d'enquêtes (gestion du temps intégrée)		Enquête italienne sur l'intégration de la deuxième génération	Comparaisons entre les données tirées d'enquête et les données administratives sur les migrations internationales au Royaume-Uni
			Études longitudinales au Royaume-Uni	
Micrecensement allemand Enquête sur la population active au Royaume-Uni				
	Enquête longitudinale canadienne sur les immigrants au Canada	Estimations des migrations internationales à long terme au Royaume-Uni sur une année glissante, tirées de l'enquête sur les voyageurs internationaux		

C. Résumé

30. Les données longitudinales nous permettent d'étudier et de comprendre les événements et les transitions qui se produisent dans le temps au cours de la vie et, chose importante, d'une génération à l'autre. Cela peut être particulièrement utile pour l'étude des migrations internationales et de l'intégration des migrants, car les données longitudinales peuvent éclairer les résultats géographiques et socioéconomiques des migrations. Ce chapitre donne un aperçu des différentes sources de données longitudinales ainsi que des exemples propres aux pays, notamment concernant l'utilisation d'enquêtes longitudinales par panel, de données administratives, y compris les registres de la population, et d'un panache d'enquêtes et de données administratives.

31. Les organismes nationaux de statistique n'auront pas tous accès à des données administratives ou aux fonds nécessaires pour mener des études longitudinales sur les migrants, de sorte qu'une approche plus pragmatique pourrait être nécessaire. Par exemple, les enquêtes transversales peuvent fournir des informations longitudinales et des perspectives

dans le temps, des questions rétrospectives pouvant être ajoutées aux enquêtes existantes. En l'absence de données longitudinales, il est également possible d'envisager de suivre des cohortes synthétiques dans le temps, ces cohortes étant toutes décrites de manière assez détaillée dans le présent chapitre.

32. Lorsque des données longitudinales sont disponibles, les difficultés et les limites connexes sont évidentes, tout comme le fait qu'elles varient selon chaque source de données. Les méthodes traditionnelles, telles que les enquêtes longitudinales par panel, présentent des avantages s'agissant du contrôle offert aux organismes de statistique, notamment en matière de chronologie, de couverture et de mesure. Cela étant, du fait que les migrants sont souvent géographiquement plus mobiles que les non-migrants et qu'ils sont généralement plus difficiles à contacter aux fins du suivi, ces sources de données présenteront souvent les défauts suivants : échantillons plus petits, phénomène d'attrition et fréquence insuffisante de la collecte. Tous ces éléments peuvent entraver l'utilisation de ces données pour l'étude de la situation à court ou à long terme des populations migrantes.

33. On observera également des limites concernant l'utilisation de données administratives telles que les registres de la population ou le processus d'intégration des données administratives pour créer un ensemble de données longitudinales. L'intégration des microdonnées, qui suppose la corrélation et le suivi des individus, constitue un défi. Le couplage des dossiers individuels dans le temps, que ce soit entre les mises à jour d'un registre de la population ou entre le registre de la population et d'autres données administratives, est souvent difficile, même lorsqu'il existe un numéro d'identification personnel. Ces questions sont décrites en détail dans le présent chapitre, mais un facteur restrictif essentiel tient au fait que ces données n'ont pas été collectées à des fins statistiques. Par conséquent, il y a une perte de contrôle sur la chronologie, la couverture de la population et l'alignement sur les concepts et les définitions de la migration. Cependant, les sources administratives peuvent généralement produire des échantillons plus grands que les enquêtes longitudinales traditionnelles, le suivi étant plus régulier et le niveau d'attrition dû à l'absence de réponse étant nul ou faible. Elles sont donc généralement bien adaptées à l'étude longitudinale de petites populations de migrants.

34. Ces défis et limites sont repris dans le chapitre 3, qui décrit les meilleures pratiques à suivre pour créer un ensemble de données longitudinales pour les statistiques migratoires et vise à traiter la question des limites dont souffrent les différentes sources de données.

III. Chapitre 3 : Comment élaborer un ensemble de données longitudinales pour les statistiques migratoires au moyen de données intégrées

A. Introduction

35. On trouvera dans le présent chapitre des indications sur la manière d'élaborer un ensemble de données longitudinales pour les statistiques migratoires à partir de données intégrées, compte tenu du Modèle générique du processus de production statistique ([Generic Statistical Business Process Model \(GSBPM\)](#)) (CEE, 2019). L'accent est mis sur les éléments propres à l'intégration des données, aux données longitudinales et aux statistiques migratoires. Le document d'orientation de la CEE sur l'intégration des données pour mesurer les migrations ([Guidance on data integration for measuring migration](#)) (CEE, 2019) devrait être utilisé en complément, notamment pour les références à l'intégration des données.

36. Chaque section du chapitre vise une phase du processus et explique comment entreprendre chaque étape à l'aide d'exemples concrets provenant de divers organismes nationaux de statistique. Des problèmes particuliers seront recensés et des solutions proposées.

37. L'élaboration d'un ensemble de données longitudinales pour les statistiques migratoires à partir de données intégrées comporte les phases suivantes :

- Conception statistique ;
- Évaluation et prétraitement des fichiers sources ;
- Intégration des données pour les données longitudinales ;
- Attribution d'identificateurs longitudinaux individuels ;
- Création de la base de données finale ;
- Diffusion des résultats.

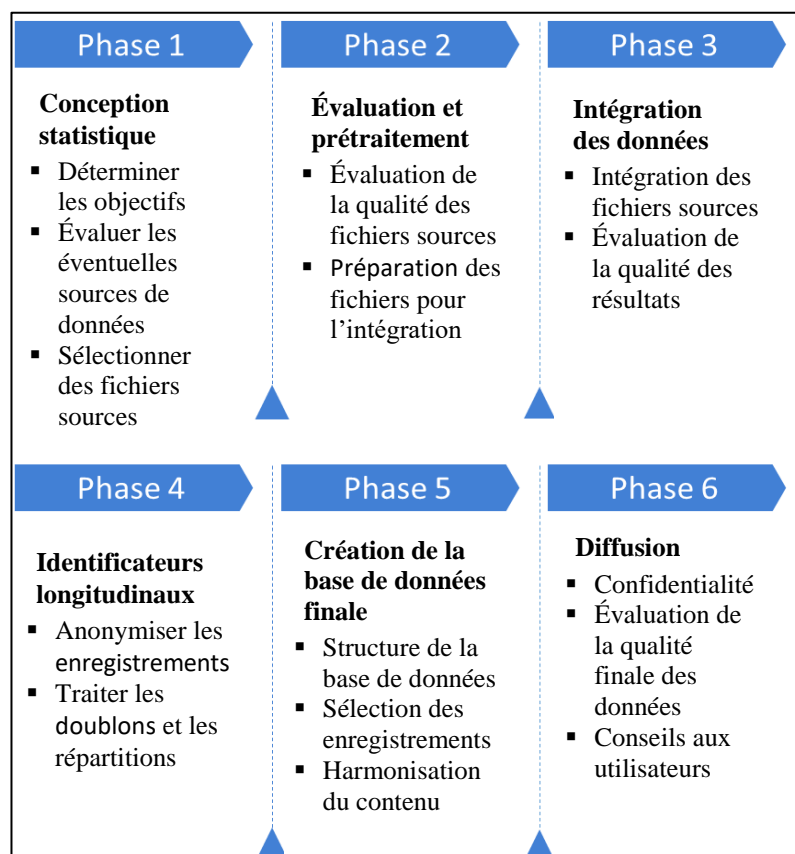
38. Des orientations sont données concernant chacune de ces phases. Elles doivent être prises en compte à la fois dans le contexte de l'élaboration d'une nouvelle base de données et dans celui de la mise à jour d'une base de données existante (par exemple pour ajouter de nouveaux résultats en intégrant des données).

39. L'ordre des phases est important. Certaines mesures peuvent toutefois être répétées selon les résultats des phases ultérieures. Par exemple, l'évaluation des fichiers sources (phase 2) peut nécessiter un réexamen de la conception statistique (phase 1). Il est également possible que les problèmes de qualité des données liés au processus d'intégration n'aient pu être décelés que durant l'évaluation finale de la base de données (phase 5). Cela peut nécessiter une nouvelle intégration des données (phase 3) ainsi qu'une répétition des phases ultérieures.

40. Il n'est pas nécessaire d'agir dans tous les cas même si toutes les étapes doivent être prises en compte. Ainsi, pour actualiser un projet existant, il n'est peut-être pas indispensable de reconsidérer la phase de conception statistique et, dans le cas de la production de données longitudinales à partir d'un système de registres, la phase d'intégration des données peut être simple.

Figure 3

Phases d'élaboration d'un ensemble de données longitudinales



B. Cadre relatif aux erreurs des données longitudinales provenant de sources de données intégrées

41. La qualité des données est un élément récurrent à prendre en considération dans toutes les phases. Il faudrait sélectionner les fichiers sources de manière qu'ils répondent au mieux aux objectifs de la conception statistique, la qualité des fichiers sources potentiels étant essentielle à la prise de cette décision. La qualité des fichiers sources est évaluée plus avant dans la deuxième phase et, si possible, améliorée à l'aide de diverses méthodes. L'intégration des données elle-même requiert une évaluation des erreurs d'intégration et, pour attribuer des identificateurs longitudinaux individuels, il faut procéder à un certain rapprochement entre les résultats de l'intégration des données. Lorsque la base de données finale est créée, des méthodes sont appliquées pour harmoniser les concepts dans le temps et préparer la base de données à l'analyse. Enfin, tous les aspects de la qualité des données doivent être étayés et des orientations adéquates doivent être données aux utilisateurs avant la diffusion des résultats.

42. Il est utile de s'inspirer du cadre relatif à l'erreur totale dans les enquêtes qui a été élaboré pour faciliter la conception statistique (Groves, 2004), (Zhang, 2012). Cela permet d'obtenir une taxonomie utile des différentes sources d'erreur potentielles, concernant principalement les données administratives. Les types d'erreurs ne sont pas tous mesurables, mais il faut les garder à l'esprit durant la conception d'un ensemble de données intégrées longitudinalement. Ce qui suit est une extension du cadre établi par Zhang (2012), qui permet de discuter plus en détail des sources d'erreur existantes lorsque l'on couple des sources administratives pour produire des ensembles de données longitudinales.

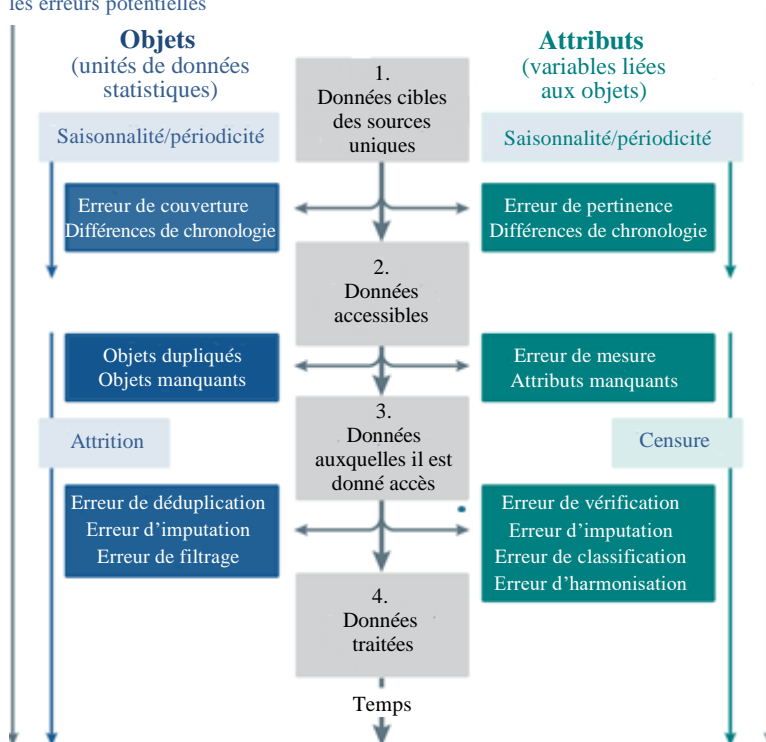
43. Le cadre, créé par le Royaume-Uni (Blackwell & Rogers, 2019), propose une approche progressive qui permet de comprendre la qualité des données. On examine d'abord des ensembles de données uniques avant d'évaluer les erreurs dans la production des ensembles de données créés en intégrant des sources multiples. Comme dans le cadre proposé par Zhang (2012) et développé ensuite par Statistics New Zealand (Statistics New Zealand, 2016), on prend en compte les erreurs liées aux objets de données (enregistrements) et à leurs attributs (variables). Toutefois, dans les données administratives, ces erreurs sont *indiquées* par les ensembles de données sources ; les attributs ne sont pas conçus comme dans une enquête et ce qui est déjà collecté est adapté en vue d'un but différent. Au cœur du cadre se trouvent des ensembles de données qui n'ont pas fait l'objet d'une conception et ils sont ce qu'ils sont.

44. Les figures 4 et 5 décrivent les cadres relatifs aux erreurs liées aux sources uniques et multiples. Les tableaux 3, 4 et 5 définissent les erreurs à chaque étape. Le cadre relatif aux sources uniques définit quatre étapes différentes dans le parcours des données : les données cibles, les données accessibles, les données auxquelles il est donné accès et les données traitées. Les erreurs sont représentées sous la forme d'une différence conceptuelle entre les données dans chacune de ces étapes. Les données cibles sont conceptuelles – les données idéales à collecter et donc les erreurs entre les données cibles et les données accessibles sont conceptuelles.

Figure 4
Cadre relatif aux erreurs liées aux sources uniques

Les améliorations et des données supplémentaires réduisent les erreurs potentielles

Facteurs externes

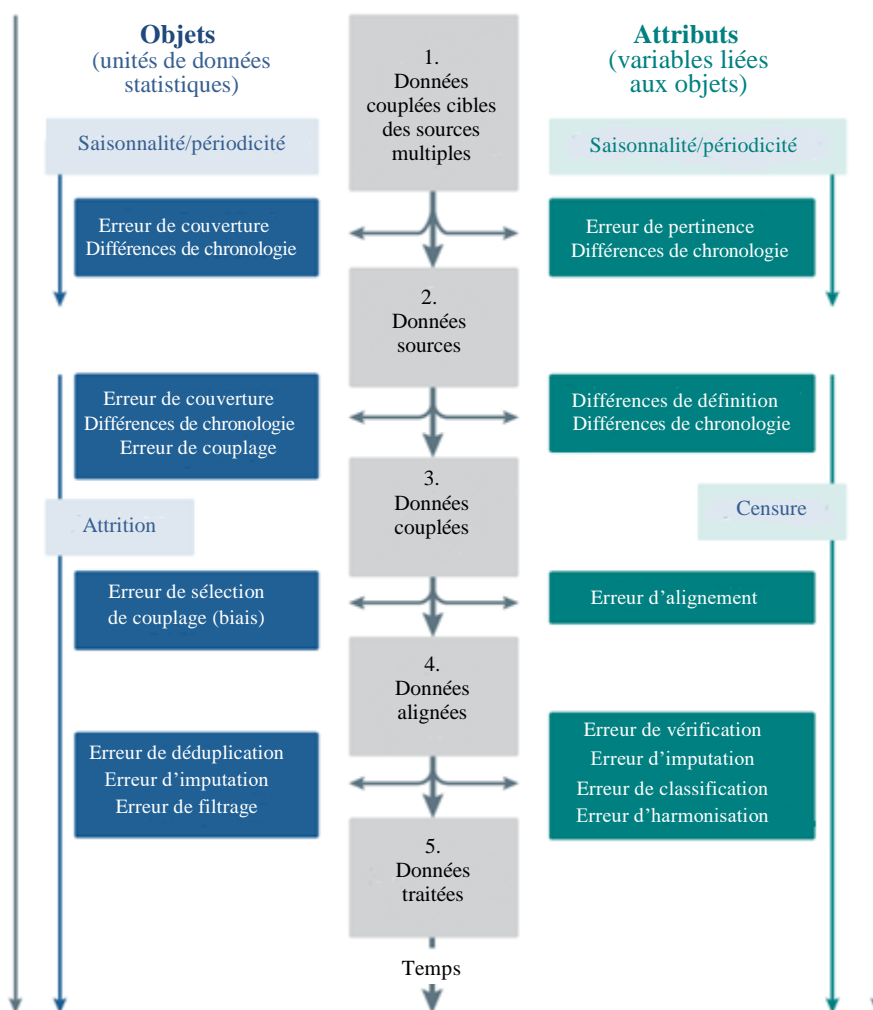


Note : Le temps est présenté verticalement. Les données cibles représentent les données idéales à collecter. Les objets (unités de données statistiques) concernent les lignes de la source de données et ce qu'elles représentent. Les attributs (variables associées aux objets) ont trait aux colonnes de la source de données et à ce qu'elles représentent.

Figure 5
Cadre relatif aux erreurs liées aux sources multiples

Les améliorations et des données supplémentaires réduisent les erreurs potentielles

Facteurs externes



45. Les erreurs sont réparties entre les objets et les attributs : les erreurs concernant les objets ont trait à l'entité à laquelle les données se rapportent, qu'il s'agisse de personnes, d'événements ou d'entreprises. Les erreurs relatives aux attributs concernent ce que l'on mesure pour les objets. Les erreurs concernant tant les objets et que les attributs peuvent influencer les uns sur les autres (représentation par une double flèche entre les deux). Les types d'erreurs ne seront pas tous applicables à chaque source de données et d'autres sources d'erreurs peuvent être décelées, notamment pour les données traitées. Dans le présent rapport, les objets concernent généralement les individus définis par des fichiers de cohortes (ou des cohortes définies par des fichiers de résultats) tandis que les attributs peuvent désigner les résultats définis par des fichiers de résultats.

46. L'erreur se propage dans le cadre, bien qu'elle n'apparaisse nécessairement pas à chaque étape. Au Royaume-Uni, ce cadre a été utilisé pour aider les chercheurs de l'Office for National Statistics à comprendre les données administratives liées longitudinalement, y compris les données du Ministère de l'intérieur relatives aux vérifications à la sortie. L'objectif est d'évaluer son utilisation pour les statistiques sur les migrations internationales. La production des données relatives aux vérifications à la sortie commence par la transmission au Ministère de l'intérieur des données concernant les voyageurs, qui sont fournies par les transporteurs commerciaux aux points d'entrée et de sortie du Royaume-Uni (Ministère de l'intérieur du Royaume Uni, 2019). Le Ministère de l'intérieur traite et gère la qualité de ces données en les couplant à d'autres données opérationnelles concernant les ressortissants étrangers soumis à l'obligation de visa.

47. Cette application a démontré que les erreurs ne se cumulent pas forcément dans les données administratives couplées. En effet, les experts du Ministère de l'intérieur cherchent à corriger les erreurs dans les données lors des divers processus. Les erreurs peuvent se cumuler au fil du temps, mais il est également possible que le fait de les compenser complique les choses. Cette compensation (par exemple l'imputation des valeurs manquantes) peut ne pas être visible dans les distributions transversales des données, mais peut s'accroître longitudinalement. Ceci est expliqué dans le contexte du cadre des sources multiples (fig. 6).

48. Le cadre intègre également l'erreur longitudinale créée par la collecte des données dans le temps – saisonnalité/périodicité, attrition et censure (tableau 4).

Tableau 1

Erreurs liées aux sources uniques

Objets¹	Attributs²
<p>Erreur de base d'échantillonnage : <i>Erreur de couverture</i> Évaluation des objets ne se trouvant pas dans les données cibles, ou impossibilité d'accéder aux objets se trouvant dans les données cibles.</p> <p><i>Différences de chronologie</i> Objets se trouvant dans les données cibles idéales qui ne sont pas accessibles en raison d'un décalage du délai d'obtention des observations.</p> <p>Erreur de sélection : <i>Objets dupliqués</i> Objets représentés plus d'une fois dans les données auxquelles il est donné accès.</p> <p><i>Objets manquants</i> Objets qui, <i>en théorie</i>, sont accessibles mais ne figurent pas dans les données auxquelles il est donné accès.</p> <p>Erreur de traitement : <i>Erreur de déduplication</i> Erreurs résultant de la déduplication d'objets dans l'ensemble de données auquel il est donné accès. Il peut s'agir à la fois de la déduplication d'objets qui sont en réalité différents (erreur « faux positif ») ou du fait de ne pas dédupliquer des objets identiques (erreur « faux négatif »).</p> <p><i>Erreur d'imputation</i> Erreurs résultant de l'imputation d'objets manquants.</p> <p><i>Erreur de filtrage</i> Erreurs résultant de la sélection ou de la désélection d'objets auquel il est donné accès concernant un ensemble de cibles idéales.</p>	<p>Erreur de pertinence : <i>Erreur de validité</i> Différence entre la mesure idéale des attributs recherchés concernant un objet et la mesure opérationnelle utilisée pour le collecter.</p> <p><i>Différences de chronologie</i> Décalage conceptuel dans la chronologie de la mesure des attributs entre les données cibles idéales et les données accessibles.</p> <p>Erreur de mesure : <i>Erreur de mesure</i> Erreurs résultant d'attributs qui ne sont pas enregistrés avec exactitude.</p> <p><i>Attributs manquants</i> Attributs manquant dans les données auxquelles il est donné accès (il peut s'agir de certains objets ou de l'ensemble des objets).</p> <p>Erreur de traitement : <i>Erreur de vérification</i> Erreurs résultant de la vérification de la valeur d'un attribut. Il peut s'agir d'une vérification faite à la suite de contrôles de validation ou d'assurance de la qualité.</p> <p><i>Erreur d'imputation</i> Erreurs résultant de l'imputation de valeurs d'attributs manquantes.</p> <p><i>Erreur de classification :</i> Erreurs résultant de la classification des valeurs en groupes ou de la détermination de nouveaux attributs.</p> <p><i>Erreur d'harmonisation :</i> Erreurs découlant de l'harmonisation des valeurs d'attributs avec un concept idéal ou cible</p>

¹ Unités de données, qui peuvent être des événements, des transactions, des personnes, des ménages, des entreprises ou d'autres enregistrements dans un ensemble de données administratives.

² Mesures ou variables qui ont été collectées et qui se rapportent aux objets ou unités de données.

Tableau 2

Erreur longitudinale –s’applique à la fois aux événements liés aux sources uniques et aux ensembles de données longitudinales multisources

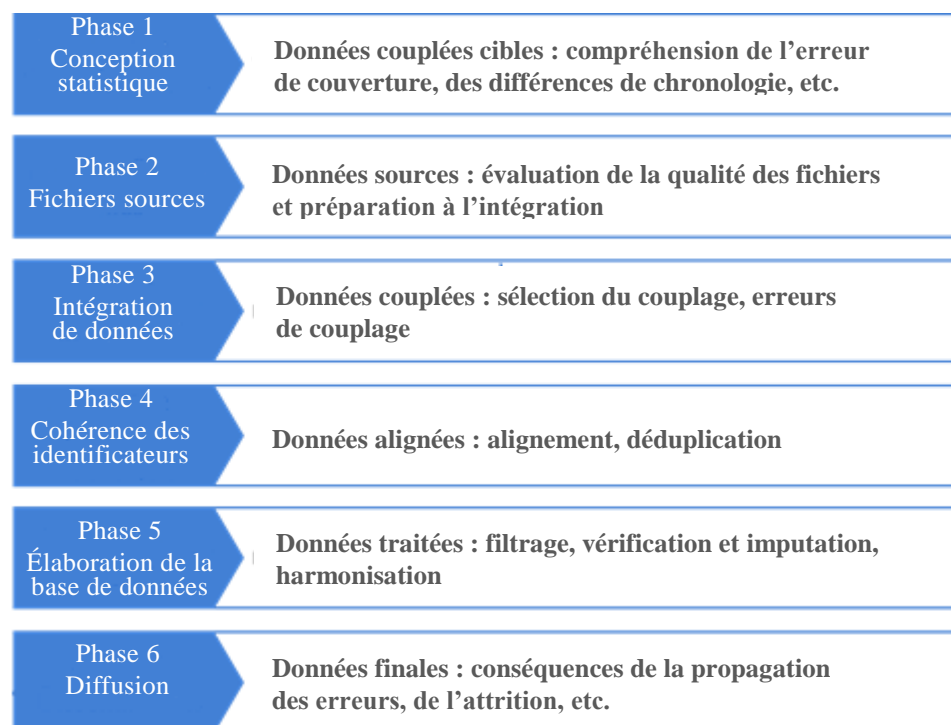
Objets¹	Attributs²
<p><i>Attrition</i> Perte d’objets ou d’unités de recherche dans le temps. Elle se produit naturellement, par un décès (ou une migration non observée). Elle peut aussi être due à un défaut de suivi, à un refus de participer, dans le cas de données tirées d’enquête, ou encore à l’absence d’informations ou à un défaut de couplage, lorsqu’il s’agit de sources administratives.</p> <p><i>Erreur de périodicité/saisonnalité</i> Les objets ne sont pas observés parce que la capture de données n’est pas assez fréquente (périodicité) ni adéquate pour rendre compte de la saisonnalité dans les données (saisonnalité).</p>	<p><i>Censure</i> Cas où la valeur d’une mesure ou d’une observation n’est que partiellement connue. Il y a une censure à droite lorsque l’objet de recherche disparaît des données avant la fin du délai d’observation ou qu’aucun événement intéressant connexe n’est observé pendant le délai d’observation. Il y a une censure à gauche lorsque l’événement intéressant s’est déjà produit avant le début du délai d’observation.</p> <p><i>Erreur de périodicité/saisonnalité</i> Les mesures des attributs dans le temps ne sont pas assez fréquentes (périodicité) ni adéquates pour rendre compte de la saisonnalité dans les données (saisonnalité).</p>

¹ Unités de données, qui peuvent être des événements, des transactions, des personnes, des ménages, des entreprises ou d’autres enregistrements dans un ensemble de données administratives.

² Mesures ou variables qui ont été collectées et qui se rapportent aux objets ou unités de données.

49. Le cadre relatif aux sources multiples (tableau 3) concerne l’intégration de sources de données multiples. Il se compose de cinq étapes : données couplées cibles, données sources, données couplées, données alignées et données traitées. Dans le contexte du présent rapport, ces étapes s’alignent sur les phases visées dans le présent chapitre, comme le montre la figure 6.

Figure 6

Étapes de l’intégration des sources de données multiples

50. Les erreurs sont représentées sous la forme d'une différence conceptuelle entre les données dans chacune de ces étapes. De nombreuses erreurs sont conceptuellement similaires qu'il s'agisse du cadre relatif aux sources multiples ou de celui relatif aux sources uniques. La principale différence tient au fait que les mesures portent sur les erreurs observées entre les ensembles de données sources et les données couplées cibles idéales (plutôt que les données cibles) ainsi que sur les erreurs observées entre les ensembles de données sources à coupler. Ces erreurs conceptuellement similaires portent le même nom dans les cadres relatifs aux sources uniques et aux sources multiples (par exemple, différences de chronologie, erreur de couverture, erreur de pertinence, erreur d'imputation, erreur de sélection et erreur de traitement).

51. Les données couplées cibles sont différentes des données cibles pour chaque source individuelle. Elles sont susceptibles de s'appliquer spécifiquement au groupe d'objets à mesurer en couplant des sources multiples. Pour la dernière étape (données traitées), le traitement peut avoir eu lieu dans le cadre relatif aux sources uniques ou aux sources multiples. L'imputation peut ainsi avoir déjà eu lieu au niveau des sources uniques ou se faire après le couplage.

Tableau 3

Erreurs liées aux sources multiples

Objets¹	Attributs²
<p><i>Erreur de base d'échantillonnage :</i> <i>Erreur de couverture</i> Observation des objets ne se trouvant pas dans les données couplées cibles, ou impossibilité d'accéder aux objets se trouvant dans les données couplées cibles.</p> <p><i>Différences de chronologie</i> Les objets ne sont pas observés en raison d'écarts conceptuels dans la chronologie de la capture des données couplées cibles et des données sources.</p> <p><i>Erreur de couverture :</i> <i>Erreur de couverture</i> Les objets ne sont pas couplés en raison des différences de couverture des objets entre les sources de données.</p> <p><i>Différences de chronologie</i> Différence entre les objets observés dans les ensembles de données sources, en raison du fait que les données sont capturées à des moments différents.</p> <p><i>Erreur de couplage</i> Erreurs résultant d'un couplage incorrect des objets (erreur « faux positif ») et du défaut de couplage des objets qui auraient dû être couplés (erreur « faux négatif »).</p> <p><i>Erreur d'identification :</i> <i>Erreur de sélection en matière de couplage (biais)</i> Erreurs résultant de la sélection d'objets couplés (ou de la désélection d'objets non couplés) en raison de biais dans le couplage ou d'une erreur dans la résolution des couplages incompatibles.</p>	<p><i>Erreur de pertinence :</i> <i>Erreur de pertinence</i> Différences entre la mesure idéale des attributs recherchés concernant un objet et les mesures opérationnelles utilisées pour la collecte dans chaque ensemble de données.</p> <p><i>Différences de chronologie</i> Écart conceptuel dans la chronologie de la mesure des attributs des données couplées cibles et des données sources.</p> <p><i>Erreur de mappage :</i> <i>Différences de définition</i> Différences entre les façons dont les attributs sont mesurés opérationnellement dans chacun des ensembles de données sources.</p> <p><i>Différences de chronologie</i> Différences entre les valeurs des attributs relatifs à un objet couplé dans les ensembles de données sources en raison du fait que les données sont capturées à des moments différents.</p> <p><i>Erreur de comparabilité :</i> <i>Erreur d'alignement</i> Erreurs résultant de l'alignement des valeurs incompatibles des attributs dans les différentes sources.</p>

Objets ¹	Attributs ²
<p><i>Erreur de traitement :</i> <i>Erreur d'imputation</i> Erreurs résultant de l'imputation d'objets manquants. <i>Erreur de filtrage</i> Erreurs résultant de la sélection ou de la désélection d'objets auquel il est donné accès dans un ensemble de cibles idéales.</p>	<p><i>Erreur de traitement :</i> <i>Erreur de vérification</i> Erreurs résultant de la vérification de la valeur d'un attribut. Il peut s'agir d'une vérification faite à la suite de contrôles de validation ou d'assurance de la qualité. <i>Erreur d'imputation</i> Erreurs résultant de l'imputation de valeurs d'attributs manquantes. <i>Erreur de classification :</i> Erreurs résultant de la classification des valeurs en groupes ou de la détermination de nouveaux attributs. <i>Erreur d'harmonisation :</i> Erreurs résultant de l'harmonisation des valeurs d'attributs avec un concept idéal ou cible.</p>
<p><i>Attrition</i> Perte d'objets ou d'unités de recherche dans le temps. Elle se produit naturellement, par un décès (ou une migration non observée). Elle peut aussi être due à un défaut de suivi, à un refus de participer, dans le cas de données tirées d'enquêtes, ou encore à l'absence d'informations ou à un défaut de couplage, lorsqu'il s'agit de sources administratives.</p>	<p><i>Censure</i> Cas où la valeur d'une mesure ou d'une observation n'est que partiellement connue. Il y a une censure à droite lorsque l'objet de recherche disparaît des données avant la fin du délai d'observation ou qu'aucun événement intéressant connexe n'est observé pendant le délai d'observation. Il y a une censure à gauche lorsque l'événement intéressant s'est déjà produit avant le début du délai d'observation.</p>
<p><i>Erreur de périodicité ou de saisonnalité</i> Les objets ne sont pas observés parce que la capture de données n'est pas assez fréquente (périodicité) ni adéquate pour rendre compte de la saisonnalité dans les données (saisonnalité).</p>	<p><i>Erreur de périodicité ou de saisonnalité</i> Les mesures des attributs dans le temps ne sont pas assez fréquentes (périodicité) ni adéquates pour rendre compte de la saisonnalité dans les données (saisonnalité).</p>

¹ Unités de données, qui peuvent être des événements, des transactions, des personnes, des ménages, des entreprises ou d'autres enregistrements dans un ensemble de données administratives.

² Mesures ou variables qui ont été collectées et qui se rapportent aux objets ou unités de données.

52. L'application du cadre relatif aux erreurs des données administratives vise à garantir une optimisation de la qualité des données dans toutes les phases du projet statistique. Dans un premier temps, il est essentiel de recenser et d'examiner les sources d'erreur pour prendre des décisions en matière de conception statistique et procéder à d'autres couplages. Pour autant, il est important de reconsidérer également les différentes sources d'erreur dans toutes les phases ultérieures. Les listes des tableaux 4 et 5 proposent un certain nombre d'indicateurs de qualité. Il n'est peut-être pas efficace pour un organisme national de statistique de mesurer chaque indicateur, mais il y a lieu de déterminer les indicateurs essentiels à inclure dans un rapport sur la qualité (voir la section **Erreur ! Source du renvoi introuvable.** dans la version intégrale des indications).

53. Une interaction existe entre les différentes sources d'erreur, par exemple, entre une erreur de couplage, une erreur de couverture et une erreur d'imputation. Les enregistrements qui contiennent des données de mauvaise qualité, peut-être à cause d'une erreur de mesure, peuvent également être plus difficiles à coupler, peut-être à cause de la qualité des identificateurs utilisés à cet effet. Une solution consiste à élaborer des méthodes sophistiquées de couplage des enregistrements pour minimiser les faux négatifs (et donc accepter plus de faux positifs) et maximiser la couverture de l'ensemble de données couplées. Il est possible cependant que les attributs qui se rapportent à ces objets soient également de mauvaise qualité et qu'ils produisent des lacunes dans les champs d'attributs ou nécessitent

une imputation. L'imputation n'est souvent pas souhaitable dans les données longitudinales, car elle peut introduire des résultats erronés. La préoccupation majeure pourrait être d'éviter l'absence de données et l'imputation.

54. Il existe également une interaction entre les erreurs liées aux objets (ou cohortes) et les erreurs liées aux attributs (ou résultats), ainsi qu'entre les ensembles de données de sources uniques et les ensembles de données de sources multiples. Les données d'attributs manquantes ou mal enregistrées peuvent influencer sur la capacité à dédupliquer ou à coupler des enregistrements dans la phase concernant les sources uniques, ce qui, à son tour, a une incidence sur la capacité à coupler les objets et crée par conséquent des erreurs dans l'ensemble de données multisources.

IV. Chapitre 4 : Diffusion de statistiques migratoires régulières à partir de sources de données longitudinales

55. On trouvera dans le présent chapitre des orientations sur la manière dont des statistiques migratoires régulières peuvent être diffusées au moyen de sources de données longitudinales. Il y sera présenté une série d'indicateurs longitudinaux qui pourraient être produits pour les statistiques migratoires. Ces indicateurs concernent notamment des sujets liés au processus migratoire lui-même ainsi que la situation socioéconomique des migrants dans le temps. Le chapitre décrit les difficultés particulières liées à chaque thème et propose des approches susceptibles d'y remédier en donnant des exemples émanant de différents pays. Les difficultés et les exemples portent en particulier sur l'utilisation de données administratives ou de sources de données intégrées.

56. S'appuyant sur la publication de la CEE intitulée « [Measuring change in the socio-economic conditions of migrants](#) » (UNECE 2015) » (Mesurer l'évolution des conditions socioéconomiques des migrants) (CEE, 2015), le chapitre propose des indicateurs sur les résultats longitudinaux, c'est-à-dire sur les résultats observés sur une période donnée.

57. Dans le présent rapport, les « indicateurs » se rapportent aux statistiques d'intérêt général pouvant être régulièrement reproduites, qui diffèrent des statistiques générées aux fins d'une recherche donnée. Les indicateurs pourraient servir à éclairer les décisions en matière de politiques ou de programmes publics, mais pourraient également être utiles à d'autres parties prenantes, notamment les organisations communautaires et les migrants eux-mêmes.

58. Dans un premier temps, les thèmes des indicateurs clefs sont énumérés pour les principaux domaines d'intérêt des données longitudinales. Plusieurs indicateurs pourraient être proposés pour chaque thème. Alors que les thèmes des indicateurs clefs peuvent se référer à des termes généraux tels que « temps écoulé avant un événement », des indicateurs bien précis pourraient inclure la proportion de migrants ayant vécu cet événement après X ans, le temps moyen écoulé avant l'événement ou d'autres mesures fondées sur le temps. L'ensemble d'indicateurs pourrait être défini pour des sous-ensembles de migrants (par exemple, les demandeurs d'asile, les résidents temporaires, etc.) et pourrait être recoupé avec des variables démographiques ou ayant trait à la situation socioéconomique (par exemple, l'âge, le sexe, le pays de naissance, l'éducation, les conditions d'admission, etc.).

59. Pour les besoins du chapitre, le terme « migrant » sera généralement utilisé pour l'ensemble des thèmes des indicateurs et des analyses. Certains indicateurs peuvent cependant ne s'appliquer qu'à certaines sous-populations. Ainsi, les migrants ne seraient pas tous concernés par les indicateurs liés aux changements de statut juridique ou de citoyenneté. Il est important de prendre en compte, le cas échéant, les populations intéressantes pour chaque ensemble d'indicateurs. Dans les exemples, les populations particulières de migrants sont clairement indiquées.

60. Chaque indicateur sera décrit et les questions liées à sa production seront examinées. Des exemples pratiques seront donnés, y compris lorsque l'indicateur pratique s'écarte de l'indicateur souhaité en raison des limites des sources de données. Le chapitre a pour but de fournir des exemples de statistiques descriptives qui peuvent être utilisées comme indicateurs longitudinaux réguliers. Il n'aborde pas les méthodes d'analyse longitudinale plus complexes

telles que l'analyse de survie ou les équations d'estimation généralisées, mais qui pourraient être utilisées pour obtenir davantage d'informations sur les résultats longitudinaux des migrants. Ces méthodes permettraient d'utiliser les ensembles d'indicateurs recensés comme variables de résultats.

61. Il est important de tenir compte des limites générales des sources de données disponibles lorsque l'on examine quels indicateurs peuvent être estimés. Le chapitre 2 présente certains des points forts et certaines limites qui caractérisent les enquêtes, les données administratives et les registres, tandis que le chapitre 3 décrit plus en détail les limites qui peuvent exister pour les sources de données intégrées.

62. L'un des défis majeurs concerne la mesure temporelle nécessaire aux indicateurs longitudinaux. Si les données ne sont pas observées en temps réel mais plutôt par segments périodiques (annuels par exemple), le temps ne peut être mesuré avec autant de précision. Une autre complication pourrait résider dans la détermination du moment où l'horloge démarre – après que le migrant arrive pour la première fois, après qu'il est devenu un résident permanent, etc. Ces questions doivent être abordées pour tous les indicateurs susmentionnés.

63. Les indicateurs longitudinaux peuvent être classés en trois grandes catégories :

- Schémas migratoires ;
- Situation socioéconomique ;
- Migrations familiales.

64. Sous chaque indicateur, un tableau indique les pays qui présentent des exemples de ces statistiques assorties des données existantes, sur une base longitudinale ou transversale. En particulier, la disponibilité des indicateurs n'est pas forcément liée à la publication régulière de ces statistiques. Des données peuvent être disponibles dans de nombreux pays, mais il est possible que ces indicateurs ne soient pas produits régulièrement mais dans le cadre d'une publication indépendante ou ponctuelle, par exemple.

Tableau 4

Indicateurs longitudinaux clefs

<i>Catégorie</i>	<i>Thème</i>	<i>Indicateur</i>	
Schémas migratoires	Durée du séjour dans le pays	Proportion de migrants toujours présents dans le pays après X temps	
		Délai moyen avant que les migrants quittent le pays	
		Durée du séjour (cumulé) depuis la première arrivée	
	Durée du séjour dans la zone géographique infranationale	Durée du séjour depuis la dernière arrivée	Durée moyenne du séjour (lorsqu'il y a plus d'un séjour par personne)
			Proportion de migrants qui restent dans la zone géographique infranationale après X temps
		Temps écoulé jusqu'au changement de	Délai moyen avant que les migrants quittent le pays
			Proportion de migrants qui restent dans la zone géographique infranationale après X temps
Temps écoulé jusqu'au changement de	Temps écoulé jusqu'au changement de	Parmi ceux qui restent dans le pays, délai moyen avant que les migrants quittent la zone géographique infranationale	
		Proportion de résidents temporaires ou de courte durée qui deviennent des résidents permanents ou de longue durée après X temps	

<i>Catégorie</i>	<i>Thème</i>	<i>Indicateur</i>
	résidence ou de statut juridique, y compris la citoyenneté	<p>Délai moyen avant que les résidents temporaires ou de courte durée deviennent des résidents permanents ou de longue durée</p> <p>Proportion de demandeurs d'asile qui sont admis à s'installer de manière permanente après X temps</p> <p>Délai moyen avant que les demandeurs d'asile soient admis à s'installer de manière permanente</p> <p>Proportion de migrants nés à l'étranger qui acquièrent la nationalité après X temps</p> <p>Délai moyen avant que les migrants nés à l'étranger acquièrent la citoyenneté nationale</p>
	Indicateurs des migrations circulaires	<p>Nombre moyen de séjours par « migrant circulaire » (à définir plus précisément, comme pour l'observation censurée, la période de temps, les durées minimales de séjour requises, etc.)</p> <p>Durée moyenne de séjour par « migrant circulaire »</p>
Situation socioéconomique	Compétences linguistiques	<p>Proportion de migrants qui sont capables de parler les langues officielles du pays après X temps</p> <p>Délai moyen avant que les migrants soient capables de parler les langues officielles du pays</p> <p>Proportion de migrants qui sont capables de parler les langues officielles du pays après X temps</p> <p>Délai moyen avant que les migrants parlent la ou les langues officielles du pays à leur domicile</p>
	Propriété du logement	<p>Proportion de migrants propriétaires de leur logement après X temps</p> <p>Délai moyen avant que les migrants soient propriétaires pour la première fois dans le nouveau pays de résidence</p> <p>Durée d'occupation du logement par les migrants (par exemple, en tant que locataire ou propriétaire) dans le temps</p> <p>Proportion de migrants vivant dans un logement subventionné ou financé par le gouvernement après X temps</p> <p>Propriété par type d'habitation (par exemple, appartement, maison jumelée, etc.) après X temps</p>
	Emploi	<p>Proportion de migrants qui sont employés après X temps</p> <p>Délai moyen avant que les migrants soient employés pour la première fois dans le nouveau pays de résidence</p> <p>Emploi à temps plein ou à temps partiel des migrants dans le temps</p> <p>Situation des migrants au regard de l'emploi dans le temps</p>

<i>Catégorie</i>	<i>Thème</i>	<i>Indicateur</i>
Situation socioéconomique	Enseignement supérieur	Proportion de migrants qui ont obtenu un certificat, un diplôme ou un titre de l'enseignement supérieur dans leur nouveau pays de résidence après X temps, par âge Fréquentation scolaire des migrants dans le temps, par âge
	Revenus	Revenu total moyen et médian après X temps Revenu moyen et médian provenant d'un emploi après X temps Proportion de migrants ayant un revenu provenant d'un emploi après X temps Proportion de migrants ayant un revenu provenant d'une activité indépendante ou d'une entreprise après X temps Proportion de migrants bénéficiant d'une aide sociale ou publique après X temps Délai avant que le revenu total moyen ou médian des migrants soit égal au revenu total moyen ou médian de la population générale Délai avant que le revenu moyen ou médian des migrants provenant d'un emploi soit égal au revenu total moyen ou médian de la population générale Proportion de migrants ayant un faible revenu après X temps
	Propriété d'une entreprise/entreprenariat	Proportion de migrants qui exercent une activité indépendante ou sont propriétaires d'une entreprise après X temps Délai moyen avant que les migrants exercent une activité indépendante ou soient propriétaires d'une entreprise Nombre de personnes employées par des migrants après X temps Tout comme les indicateurs relatifs à la situation au regard de l'emploi, les registres fiscaux peuvent fournir une indication indirecte sur l'entrepreneuriat en mentionnant la proportion de migrants touchant un revenu provenant d'une activité indépendante ou d'une entreprise. Sinon, il est possible d'utiliser les registres administratifs sur les propriétaires d'entreprises. Il peut être plus complexe de corréler les données sur les migrants entrepreneurs avec les informations se rapportant à leurs employés. Toutefois, cette analyse est également possible s'il existe des registres administratifs reliant les travailleurs aux entreprises et les propriétaires aux entreprises. La Base de données sur la dynamique canadienne entre employeurs et employés est un cadre de couplage fondé sur 12 fichiers sources

<i>Catégorie</i>	<i>Thème</i>	<i>Indicateur</i>
		administratifs, dont la base de données longitudinales sur les immigrants (BDIM). Elle comprend les caractéristiques des individus, des entreprises et des propriétaires d'entreprises. Comme il s'agit d'un cadre de couplage, les utilisateurs doivent intégrer les différentes composantes pour répondre à leurs besoins d'analyse en utilisant des identifiants uniques anonymisés pour chaque fichier. La Base de données sur la dynamique canadienne entre employeurs et employés permet d'analyser les entreprises et les individus dans le temps en répondant à des questions concernant, par exemple, la répartition initiale dans les entreprises et la croissance des gains, les entreprises détenues par des immigrants et les trajectoires de ces entreprises.
	Santé	Proportion de migrants qui sont enregistrés auprès d'un médecin dans le système de santé du pays d'accueil après X temps (ou avant l'enregistrement) Proportion de migrants qui sont hospitalisés dans les X années qui suivent leur arrivée dans le pays d'accueil de résidence, globalement ou pour diverses affections particulières, par exemple de nature chronique ou infectieuse (ou durée de la première hospitalisation) Proportion de migrants décédés depuis leur arrivée dans le pays d'accueil pour diverses maladies (après X temps) (temps écoulé avant le décès)
Migrations familiales		Délai moyen entre l'arrivée du premier membre de la famille dans le pays et l'arrivée du dernier membre de la famille (c'est-à-dire le délai de regroupement familial) Situation économique des familles de migrants après X temps dans le pays

V. Chapitre 5 : Conclusions et recommandations

A. Conclusions

65. Étudier les migrations revient à étudier les changements : changement de résidence, changement de statut juridique et changement de situation socioéconomique. En outre, les thèmes connexes de l'intégration et de l'établissement sont des processus et non des états. Les résultats pour les migrants peuvent apparaître à court ou à long terme.

66. Afin de produire des statistiques sur les migrations et les sujets connexes, il faut tirer parti des données mesurant les changements. Les données longitudinales permettent de comprendre les changements au niveau individuel. Au-delà de la simple observation des changements de résidence ou de situation des individus, le suivi de ces derniers dans le temps permet de mieux comprendre les facteurs qui peuvent influencer sur les changements. Ces facteurs peuvent agir au niveau global (changements de politique et mutations économiques par exemple) ou au niveau individuel (obtention d'un enseignement ou d'une formation et acquisition de la citoyenneté, par exemple).

67. Les chercheurs et les statisticiens officiels ont de plus en plus recours à des méthodes fondées sur des approches longitudinales pour mieux comprendre les flux migratoires et la dynamique de l'intégration dans les pays d'accueil. Les approches longitudinales conviennent le mieux pour mesurer l'évolution de l'écart entre les immigrants et la population autochtone durant le temps écoulé depuis l'immigration, par exemple en ce qui concerne les salaires ou le chômage.

68. Les données administratives étant plus disponibles, les pays peuvent désormais mieux exploiter des techniques telles que l'intégration de données pour pouvoir élaborer des données longitudinales détaillées aux fins des statistiques migratoires. Les analyses des données administratives intégrées peuvent répondre à des questions qui nécessitent des échantillons de grande taille comportant des données abondantes et détaillées sur des populations difficiles à atteindre et peuvent produire des informations d'un degré élevé de validité et d'applicabilité externe pour l'élaboration des politiques. Cette approche permet une observation longitudinale de la situation concernant la migration, l'intégration ou l'établissement sans augmenter la charge que représente la collecte de données dans le cadre d'une enquête.

69. Les présentes indications montrent comment élaborer et utiliser des données longitudinales pour les statistiques relatives aux migrations internationales. Le chapitre 2 donne un aperçu des types de données allant de celles tirées des enquêtes par panel à celles des registres de la population. Le chapitre 3 explique comment développer des sources de données longitudinales à partir de données intégrées. Enfin, le chapitre 4 propose des indicateurs longitudinaux clefs liés à l'étude des migrations internationales ainsi que les bonnes pratiques à diffuser.

70. Bien que les données longitudinales offrent une occasion exceptionnelle et importante d'étudier la migration, l'intégration et l'établissement en tant que processus, les présentes indications font apparaître les difficultés liées à ces données et leurs limites. Les difficultés varient selon le type de source de données, comme le décrit le chapitre 2.

71. Les méthodes traditionnelles, telles que les enquêtes longitudinales par panel, présentent des avantages s'agissant du contrôle offert aux organismes de statistique en matière de couverture et de mesure. Cependant, pour des raisons opérationnelles, ces sources de données ont tendance à présenter les défauts suivants : échantillons plus petits, attrition et fréquence insuffisante de la collecte des données. Ces limites peuvent entraver leur utilisation pour l'étude de la situation à court ou à long terme des populations peu nombreuses (par exemple, des sous-ensembles de migrants).

72. Par ailleurs, il existe également des limites concernant l'utilisation des données administratives intégrées. En particulier, ces données n'ont pas été collectées à des fins statistiques, d'où une perte de contrôle sur la couverture et les mesures. Le processus d'intégration peut en outre introduire des erreurs, comme nous l'avons vu au chapitre 3. Cela étant, ces méthodes produisent souvent des échantillons de grande taille, le suivi étant plus régulier et le niveau d'attrition dû à l'absence de réponse étant nul ou faible. Elles sont donc généralement bien adaptées à l'étude longitudinale de petites populations de migrants.

73. Des méthodes permettent d'atténuer, de traiter ou de contourner les limites des sources de données afin que ce qui est diffusé serve toujours les objectifs définis lors de la conception statistique. Le chapitre 3 décrit diverses pratiques optimales à suivre tout au long de l'élaboration des sources de données, tandis que le chapitre 4 montre comment des indicateurs précieux peuvent être produits malgré les limites de la base de données obtenue.

74. Il est important non seulement de ne pas ignorer les difficultés, mais aussi de ne pas se laisser entraver par elles. Il est essentiel de savoir si des statistiques particulières sont adaptées à l'objectif visé. L'intégration de données aux fins des statistiques longitudinales sur les migrations posera toujours des problèmes, mais elle ouvre également des possibilités d'analyse inédites, qui mettront en lumière les résultats longitudinaux détaillés des migrations et leurs effets sur les pays d'accueil. Il est possible que ces sources de données ne donnent pas toujours des estimations transversales ponctuelles parfaites, mais elles offrent quelque chose de plus riche – une histoire.

75. Les exemples figurant dans le chapitre 4 montrent bien la puissance des données longitudinales en ce qui concerne les statistiques migratoires. En fin de compte, il faudrait venir à bout des difficultés au vu des avantages que présentent les données longitudinales pour les statistiques migratoires.

B. Recommandations

76. Pour mieux comprendre les migrations et les processus d'intégration et d'établissement en particulier, il est recommandé d'utiliser davantage les données longitudinales pour les statistiques migratoires.

77. Pour élaborer un ensemble de données longitudinales, il est recommandé de suivre chacune des six phases décrites dans le chapitre 3 dans l'ordre indiqué. En fonction des résultats, il peut être nécessaire de revenir à l'une des phases précédentes, auquel cas les phases ultérieures devront être répétées. Les phases sont les suivantes :

- Conception statistique ;
- Évaluation et prétraitement des fichiers sources ;
- Intégration des données pour les données longitudinales ;
- Attribution d'identificateurs longitudinaux individuels ;
- Création de la base de données finale ;
- Diffusion des résultats.

78. Il est recommandé aux pays de produire régulièrement les indicateurs longitudinaux destinés aux statistiques migratoires décrits dans le chapitre 4.

C. Pistes de travail

79. Le présent rapport énumère un large éventail de thèmes liés aux migrations auxquels les données longitudinales peuvent être utiles. Certains thèmes ont été abordés plus en détail, notamment dans le chapitre 4, au moyen des indicateurs proposés. D'autres ont été cependant considérés comme ne relevant pas du champ d'application des présentes indications.

80. Il serait nécessaire en particulier d'étudier plus avant le mode de mesure des migrations familiales. En raison de sa complexité, ce thème devrait peut-être faire l'objet d'un examen plus ciblé, notamment une étude de sous-thèmes tels que les suivants :

- Schémas migratoires des familles ;
- Familles de migrants en tant qu'unités d'analyse (par exemple, l'établissement et l'intégration des familles de migrants dans le temps) ;
- Regroupements familiaux et membres vivant à l'étranger ;
- Milieu et mode de vie des migrants dans le pays d'accueil (par exemple, ménages multifamiliaux, ménages multigénérationnels).

81. Un autre thème méritant une plus grande attention est celui des schémas d'établissement infranationaux dont il est question dans la section 4.1.1.2 dans la version intégrale des indications – Durée du séjour dans la zone géographique infranationale. Cependant, de plus en plus les pays s'intéressent à la composition des quartiers et encouragent les migrants à s'installer dans des lieux non traditionnels (par exemple pour des raisons économiques). Il peut être important pour les parties prenantes telles que les prestataires de services et les écoles de comprendre les configurations géographiques de faible niveau. Pour autant, comme nous l'avons vu tout au long des présentes indications, en particulier au chapitre 4, ce type d'étude engendre des défis sans précédent. En particulier, les frontières fluctuent dans le temps et parfois un déplacement sur une courte distance pourrait être considéré comme un changement de zone géographique alors qu'un déplacement sur une longue distance pourrait être perçu comme restant dans la même zone géographique.

82. Une autre nouvelle question est celle de savoir comment les organismes de statistique diffusent les résultats. On trouvera quelques approches innovantes au chapitre 4. De nouveaux outils de diffusion étant disponibles et les données à diffuser étant plus complexes, les organismes de statistique sont incités à se pencher de nouveau sur la question de savoir comment présenter les résultats à divers publics. Il s'agit là d'un défi particulièrement important en ce qui concerne les statistiques migratoires. En raison de la récente augmentation des migrations internationales, on observe dans divers pays de nombreux discours publics qui exigent des informations sur les schémas migratoires et les effets des migrations sur les individus, les familles et les sociétés. Les statistiques migratoires étant complexes de par leur nature, les résultats statistiques risquent également d'être mal interprétés. Des orientations supplémentaires sur la manière de présenter les résultats à divers publics tout en respectant et en reconnaissant les limites quant à la qualité des données pourraient aider les organismes nationaux de statistique à relever ce nouveau défi.
