



**Conseil économique
et social**

Distr.
GÉNÉRALE

ECE/CES/2009/3
27 mars 2009

FRANÇAIS
Original: ANGLAIS

COMMISSION ÉCONOMIQUE POUR L'EUROPE

COMMISSION DE STATISTIQUE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

Cinquante-septième réunion plénière
Genève, 8-10 juin 2009
Point 6 de l'ordre du jour provisoire

**PRINCIPES RELATIFS À LA PROTECTION DE LA CONFIDENTIALITÉ ET
AU RESPECT DE LA VIE PRIVÉE DANS LE CADRE DE L'INTÉGRATION
DES DONNÉES STATISTIQUES**

**PRINCIPES ET LIGNES DIRECTRICES RELATIFS À LA CONFIDENTIALITÉ
DANS LE CONTEXTE DE L'INTÉGRATION DES DONNÉES RÉALISÉE
À DES FINS STATISTIQUES OU AUX FINS DE TRAVAUX
DE RECHERCHE CONNEXES**

**Note de l'Équipe spéciale de la confidentialité et du caractère privé
des statistiques intégrées**

Résumé

À sa réunion plénière de 2006, la Conférence a recommandé de créer une équipe spéciale chargée d'examiner les aspects de la confidentialité et du caractère privé des ensembles intégrés de données et d'élaborer des principes communs (ECE/CES/70). L'Équipe spéciale de la confidentialité et du caractère privé des statistiques intégrées a été créée en février 2007. La présente note renferme un projet de principes et de pratiques élaboré par l'Équipe spéciale. Elle a pour objet de créer un cadre commun en vue d'évaluer et d'atténuer les aspects législatifs et les autres aspects de confidentialité liés à la création et l'utilisation d'ensembles intégrés de données aux fins de statistique et de recherches connexes.

Les Principes et Lignes directrices sur les aspects de confidentialité des statistiques intégrées réalisées aux fins de statistique ou de recherches connexes sont soumis pour approbation à la Réunion plénière de la Conférence de 2009, conformément à la Procédure régissant l'adoption de produits et de recommandations par la Conférence des statisticiens européens (ECE/CES/2006/37/Rev.1).

I. INTRODUCTION

1. À sa huitième réunion, la Commission économique des Nations Unies pour l'Europe de l'ONU a adopté sa décision C(47) relative aux Principes fondamentaux de la statistique officielle dans la région de la Commission économique pour l'Europe, en date du 15 avril 1992 (voir <http://www.unece.org/stats/documents/e/1992/32.e.pdf>). Le Principe fondamental 6 dispose que «les données individuelles collectées pour l'établissement des statistiques par les organismes qui en ont la responsabilité, qu'elles concernent des personnes physiques ou des personnes morales, doivent être strictement confidentielles et ne doivent être utilisées qu'à des fins statistiques».

2. L'objet de l'intégration des données est d'intégrer des données concernant des unités statistiques provenant de différentes sources administratives et/ou d'enquêtes, afin de constituer de nouveaux ensembles statistiques susceptibles d'être publiés en tant que tels. En outre, ces ensembles intégrés peuvent être utilisés dans le cadre d'un large éventail d'études économiques et sociales, qui ne pourraient être réalisées à partir des seules sources de données traditionnelles.

3. Ces principes et Lignes directrices s'appliquent aux travaux d'intégration de données menés au sein des services nationaux de statistique. Dans certains cas, les organisations internationales de statistique combinent des ensembles de microdonnées provenant de différents pays, mais étant donné qu'il est peu probable que les mêmes unités statistiques se retrouvent dans plusieurs fichiers statistiques nationaux, on ne parle pas, dans ce cas, d'intégration des données.

4. Dans le contexte de ces principes et Lignes directrices, l'emploi d'autres sources lors des processus de validation et d'imputation des données provenant d'une source unique n'est pas considéré comme une intégration de données, même si on retrouve parfois des problèmes du même ordre. Deux éditions de la même enquête sont traitées comme une seule et même source.

5. Les définitions ci-après sont fournies dans le but de faciliter l'emploi des Principes et des Lignes directrices:

- a) Microdonnée composite: unité résultant de l'intégration de données;
- b) Confidentialité: obligation pour le fournisseur de l'information de garantir le secret de l'information qu'il fournit;
- c) Intégration des données: processus consistant à combiner les données provenant de deux sources ou davantage dans le but de produire de nouveaux ensembles;
- d) Correspondance des données: établissement du lien entre microdonnées provenant de différentes sources, à partir de caractéristiques communes présentes dans toutes ces sources;
- e) Fournisseur de données: organisation qui produit des données ou des métadonnées. Aux fins des Principes et des Lignes directrices, ce terme comprend les fournisseurs de fichiers de données provenant de sources statistiques et non statistiques, mais il exclut les répondants qui prennent part à des enquêtes;

- f) Personnes physiques ou morales: particuliers et entités juridiques reconnues aux termes de la législation nationale;
 - g) Statistique officielle: toute activité statistique effectuée dans le cadre d'un système national de statistique ou du programme statistique d'une organisation intergouvernementale;
 - h) Vie privée: le droit de toute personne de garder secrètes sa vie personnelle et ses relations, et l'obligation pour le détenteur de l'information de préserver ce secret;
 - i) Activité statistique: collecte, stockage, traitement et diffusion de statistiques;
 - j) Activités statistiques: utilisation de données en conformité avec les Principes fondamentaux relatifs à la statistique officielle, dans le cadre d'une ou de plusieurs phases du processus statistique, de façon à alimenter la production de statistiques officielles.
6. L'intégration des données peut comporter des correspondances exactes, des correspondances fondées sur la probabilité et/ou des correspondances statistiques. Les ensembles intégrés de statistiques peuvent offrir, notamment, les avantages suivants:
- a) Production de nouvelles statistiques ou amélioration des statistiques existantes;
 - b) Production de statistiques ventilées plus finement, pour des mesures portant sur des informations existantes;
 - c) Possibilité de mener des recherches à partir de microdonnées composites englobant un éventail plus large pour un nombre plus important d'unités que ce que permet une source unique;
 - d) Possibilité d'améliorer ou de valider des sources de données existantes;
 - e) Possibilité de réduire la charge de travail des répondants.
7. Les principes et lignes directrices reproduits ci-après développent le Principe fondamental 6 et définissent un cadre commun permettant d'évaluer et d'atténuer l'impact législatif et de préserver la confidentialité dans le contexte de la création et de l'utilisation d'ensembles intégrés aux fins de recherches et de statistiques. Ils affirment, en particulier, que les principes fondamentaux qui régissent la statistique officielle s'appliquent de la même façon aux ensembles intégrés qu'à toute autre source de statistiques officielles.
8. Les principes reposent sur le postulat selon lequel certains gouvernements ont pour politique de chercher d'abord à produire des statistiques officielles à partir des sources administratives de données (intégrées ou non), telles que les registres, et de ne réaliser d'enquêtes que lorsque des variables essentielles sont manquantes ou de trop mauvaise qualité dans les sources administratives. Dans les pays concernés, l'intégration d'ensembles statistiques fait partie intégrante de l'activité des services nationaux de statistique. Ces pays disposent déjà d'un cadre législatif solide et de règles claires en matière de protection de la confidentialité des données personnelles et des données des entreprises, que les données soient ou non intégrées à partir de sources différentes.

9. Cependant, nombreux sont les pays pour lesquels l'idée d'intégrer des données pour produire des microdonnées composites à partir de sources différentes aux fins de statistique ou de recherches connexes est relativement nouvelle. Les principes et les lignes directrices reproduits ci-après, de même que l'étude de cas jointe en annexe, ont pour objet de soumettre cette activité à un cadre propre à en rendre l'application plus claire et cohérente.

10. Parallèlement aux principes et lignes directrices énoncés ci-après, deux autres notions ont été définies. Elles sont susceptibles d'acquiescer de l'importance dans certaines circonstances, particulièrement pour les pays qui ne pratiquent que depuis peu l'intégration des données aux fins de statistique officielle. Premièrement, l'intégration des données ne doit pas être réalisée si elle comporte une menace significative pour l'intégrité des données collectées auprès des différentes sources. Deuxièmement, l'intégration de données aux fins de recherches ne doit être envisagée que lorsque le processus d'approbation permet de démontrer qu'il en va de l'intérêt public. On part du principe que les activités de statistique officielle serviront toujours l'intérêt public si elles obéissent aux principes fondamentaux qui les régissent. Ces notions sont évoquées ici dans un souci d'exhaustivité, mais elles ne sont pas acceptées de façon courante et n'ont donc pas le même statut que les principes énoncés ci-après.

II. PRINCIPES ET LIGNES DIRECTRICES RELATIFS À LA CONFIDENTIALITÉ DES STATISTIQUES INTÉGRÉES

A. Principe 1

L'intégration de données ne doit être réalisée que par les services nationaux de statistique, uniquement à des fins de recherche et de statistiques, et à condition que la publication de données susceptibles de révéler l'identité d'une personne physique (et, si la législation le prévoit, une personne morale telle qu'une entreprise) soit régie par la loi.

11. Lignes directrices:

a) Le principe énoncé ci-dessus doit être inscrit soit dans la législation relative aux statistiques, soit dans la législation relative à la protection des données, soit dans les deux, et les gouvernements doivent s'y conformer scrupuleusement;

b) En cas d'absence d'une protection légale explicite, les services nationaux de statistique doivent s'abstenir de procéder à l'intégration de données relatives à des personnes physiques et morales;

c) Sauf disposition législative contraire, l'utilisation par les services nationaux de statistique, aux fins de statistiques ou de recherche, de données administratives ou statistiques en possession des départements administratifs, ne constitue pas une atteinte à la vie privée d'une personne physique ou morale.

B. Principe 2

Les services nationaux de statistique ne doivent entreprendre d'activités d'intégration de données que dans le cadre de leur mandat, qui est de produire des statistiques officielles, et au terme d'une procédure d'approbation normalisée (par exemple, un cahier des charges).

12. Lignes directrices:

a) Si le mandat d'un service national de statistique va au-delà de la production de statistiques aux fins de statistiques et de recherches connexes, consistant, par exemple, à utiliser des données à des fins administratives ou réglementaires en rapport avec les personnes physiques, le service concerné doit s'abstenir de toute activité d'intégration de données aux fins de statistique ou de recherches connexes se rapportant à ces mêmes unités, sauf si la loi et/ou une autorité indépendante du service l'autorise expressément;

b) Avant d'entreprendre toute nouvelle enquête statistique, il convient de se demander si l'intégration des données permettrait de réduire les coûts, d'améliorer la qualité ou de réduire la charge de travail des répondants;

c) Une procédure d'approbation normalisée, qui pourrait, par exemple, prendre la forme d'un cahier des charges officiel, doit être appliquée pour toute nouvelle proposition d'activité d'intégration de données. Un exemple de cahier des charges est présenté dans les grandes lignes en annexe, mais il appartient à chaque pays de trouver son propre mode d'approbation des projets d'intégration de données. Cette procédure doit décrire la façon dont le travail d'intégration permettra de produire des statistiques, d'améliorer les statistiques existantes ou de contribuer aux recherches connexes.

C. Principe 3

Les avantages que comporte un projet d'intégration de données pour le public doivent être suffisants pour surpasser les inconvénients que l'utilisation de ces données pourrait avoir du point de vue de la vie privée et de la confidentialité et/ou des risques pour l'intégrité du système des statistiques officielles.

13. Lignes directrices:

a) L'intégration des données doit s'opérer dans un environnement sûr et ne doit comporter aucun risque pour l'intégrité du système des statistiques officielles;

b) À moins que la législation n'en dispose autrement ou que la procédure d'approbation normalisée prévoit d'autres modalités, tous les éléments d'identification directe associés aux données doivent être retirés dès que possible après achèvement de la procédure d'intégration;

c) Le cas échéant, les organes chargés de veiller à ce que les avantages, le souci de confidentialité et les risques pour l'intégrité du système des statistiques officielles soient dûment identifiés et pris en considération par les services nationaux de statistique dans le cadre de la procédure d'approbation normalisée doivent être consultés. La liste des avantages doit notamment faire état des avantages attendus de la conservation éventuelle de l'ensemble intégré de données envisagée à long terme ou, au contraire, de son expansion au fil du temps;

d) Dans certains pays, la législation impose que la procédure d'approbation normalisée comporte une évaluation de l'impact sur la confidentialité;

e) Le consentement des fournisseurs de données doit être sollicité lorsque cela est raisonnable et réalisable.

D. Principe 4

Il convient de ne pas intégrer de données dès lors qu'un engagement spécifique dans ce sens a été pris vis-à-vis des répondants.

14. Lignes directrices:

La procédure d'approbation normalisée (le cahier des charges par exemple) doit donner la possibilité d'étudier les engagements éventuels pris vis-à-vis des répondants concernant les cas dans lesquels leurs données peuvent être utilisées. Le Directeur du service national de statistique ne doit pas approuver un projet d'intégration de données si un des éléments de ce projet est incompatible avec de tels engagements.

E. Principe 5

Les données intégrées ne doivent être utilisées que pour des activités statistiques ou de recherche approuvées, et toute variation importante par rapport à l'activité initialement approuvée doit faire l'objet d'une nouvelle procédure d'approbation normalisée.

15. Lignes directrices:

Sauf disposition législative contraire, il y a lieu de solliciter une nouvelle approbation dans les cas suivants:

a) Les sources (ensemble de données) utilisées dans le cadre du processus d'intégration évoluent de façon importante, ou il est proposé d'ajouter une nouvelle source au processus d'intégration;

b) Il est proposé d'augmenter sensiblement le nombre de caractéristiques prises en compte aux fins de l'intégration;

c) Il est proposé d'augmenter sensiblement le nombre d'unités prises en compte aux fins de l'intégration (par exemple, le passage de quelques-uns à l'ensemble des secteurs économiques);

d) La méthode d'intégration évolue (par exemple, d'une correspondance statistique à une correspondance exacte);

e) Il est proposé d'utiliser l'ensemble intégré ainsi obtenu aux fins d'une autre activité statistique officielle ou dans le cadre d'un projet de recherche soumis par un chercheur indépendant, non prévus lors de la procédure initiale d'approbation normalisée.

F. Principe 6

Le nombre d'unités et de variables de données prises en compte dans le cadre des ensembles intégrés ne doit pas excéder le nombre nécessaire à la réalisation de l'activité approuvée.

16. Lignes directrices:

a) On entend par «activité approuvée» toute activité approuvée dans le cadre du cahier des charges spécifique à l'intégration de données. Seules les variables nécessaires à ces activités doivent être prises en compte aux fins du projet d'intégration de données approuvé;

b) Le nombre d'unités prises en compte doit correspondre au minimum requis par l'activité approuvée (il faut, par exemple, envisager la possibilité de ne prendre en compte qu'un échantillon à partir d'une source à couverture intégrale).

G. Principe 7

Les services nationaux de statistique doivent mener les projets d'intégration de données en toute ouverture et en toute transparence.

17. Lignes directrices:

a) La politique des services nationaux de statistique en matière d'intégration de données et un aperçu des projets d'intégration de données en cours doivent être publiés;

b) Les principaux résultats statistiques de tous les projets d'intégration de données doivent pouvoir être consultés par le public. Lorsque des projets d'intégration de données sont utilisés pour améliorer la production de statistiques officielles (la qualité, par exemple), la publication de ces statistiques officielles s'inscrit dans le cadre de cette amélioration;

c) Sauf disposition législative contraire, les administrations doivent, s'il y a lieu et si possible, faire savoir aux répondants que leurs données sont susceptibles d'être utilisées dans le cadre d'activités de statistique ou de recherche.

H. Principe 8

L'accès aux données composites obtenues par intégration, à l'exclusion des identifiants, doit, d'une façon générale, être restreint au personnel habilité du service national de statistique. Toute proposition visant à octroyer l'accès à ces données à une personne extérieure au service doit être juridiquement fondée et compatible avec les objectifs fixés dans le cadre de la procédure d'approbation normalisée. Toute personne ainsi autorisée doit apporter la garantie juridique, institutionnelle et logistique qu'elle utilisera les données sans déroger à la proposition qui a reçu l'approbation et qu'aucune personne non autorisée ne pourra y accéder.

18. Lignes directrices:

a) Les microdonnées composites résultant de l'intégration réalisée par le service national de statistique peuvent être utilisées à des fins de statistique ou de recherche connexes par d'autres producteurs de statistiques officielles au sein du même système national de statistique ou, si la législation nationale en dispose ainsi, dans le cadre d'un système statistique national, pour autant qu'un cahier des charges conforme à la Ligne directrice 2 b) ait reçu l'approbation du service national de statistique. Une telle approbation doit notamment indiquer si l'activité visée est totalement dissociée, du point de vue organisationnel, de toute activité de collecte et de traitement de données à des fins administratives;

b) Le service national de statistique ne doit communiquer au fournisseur de données aucune information concernant quelque variable que ce soit issue d'un ensemble intégré et susceptible de l'assister dans le cadre d'une activité d'administration ou de réglementation;

c) Si leur bonne foi est avérée, les chercheurs extérieurs peuvent bénéficier d'un accès aux microdonnées issues d'ensembles intégrés, conformément aux Lignes directrices de la Conférence des statisticiens européens pour «assurer la confidentialité des données statistiques et l'accès aux microdonnées», pour autant qu'un cahier des charges ait reçu l'approbation du service national de statistique;

d) Les obligations incombant au bénéficiaire doivent faire l'objet d'un contrat, et toute atteinte aux règles de confidentialité de la part de celui-ci est punissable d'une sanction stipulée dans la législation et exécutable non seulement à l'encontre du bénéficiaire lui-même, mais aussi, le cas échéant, de l'institution qui l'a parrainé.

ANNEXE

[ENGLISH ONLY]

EXAMPLE OF A BUSINESS CASE OUTLINE

1. Principle 2 provides that a standard approval process should be followed for any new data integration proposal and suggests that this may take the form of a formal business case. If a business case approach is to be used, it is proposed that it should cover the following topic areas:

A. Purpose(s)

2. The business case should describe the purposes for which the integrated data will be used.

B. Benefit to Official Statistics

3. The business case should describe how the proposed project will produce or improve official statistics. Improving official statistics could involve improvements to the accuracy, reliability, relevance, timeliness, consistency and coverage of the statistics or to the concepts, definitions or methods used to produce the statistics.

C. Other Benefits

4. The business case should describe who else will benefit, and how they will do so, from the project.

D. Retention

5. The business case should state how long the integrated dataset needs to be retained to support the purposes for use. The retention may be subject to periodic reviews.

E. Data sources

6. The proposal should describe what sources of data will be used for data integration. This should list the proposed source agencies and describe in general terms the data to be received from each agency.

7. Any risks to integrity of the data sources should be listed. Any implications of the legislation that source data was collected under to the data integration project should be listed.

F. Alternatives

8. It should be explained why data integration is preferable to any feasible alternative in terms of cost, quality or minimising compliance burden.

G. Stakeholders

9. The business case should list all identified key stakeholders (both internal and external) in the data integration project and the results of any consultation with them.

H. Retention of names and addresses

10. If the data integration project needs to retain personal names and addresses for linking then this should be stated along with how long retention of these details is needed.

I. Frequency of Reviews

11. The business case should specify the frequency with which reviews of the data integration will be held.

J. Privacy Impact Assessment

12. A privacy impact assessment should be completed unless a country's legislative and/or relevant NSO policy provides an exemption. It should also be noted that although privacy generally relates to natural persons, it may also relate to a legal person in the case of some businesses or industries. For example, certain unincorporated businesses, such as farms, may generate privacy considerations in some countries.
