



**Conseil économique
et social**

Distr.
GÉNÉRALE

ECE/CES/2006/6
30 mars 2006

FRANÇAIS
Original: ANGLAIS

COMMISSION ÉCONOMIQUE POUR L'EUROPE

COMMISSION DE STATISTIQUE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

Cinquante-quatrième réunion plénière

Paris, 13-15 juin 2006

Point 5 de l'ordre du jour provisoire

**LIGNES DIRECTRICES ET PRINCIPES DE BASE POUR ASSURER
LA CONFIDENTIALITÉ DES DONNÉES STATISTIQUES
ET L'ACCÈS AUX MICRODONNÉES**

**Principes et lignes directrices correspondant à une bonne
pratique – lignes directrices provisoires¹**

**Document présenté par l'Équipe spéciale chargée de la gestion
de la confidentialité et de l'accès aux microdonnées**

I. NOTE

1. Les présentes Lignes directrices ont été élaborées à la demande de la Conférence des statisticiens européens par une Équipe spéciale constituée à cet effet sous la direction du Bureau de la Conférence. M. Dennis Trewin (Bureau australien de statistique) a présidé l'Équipe spéciale. Le Bureau de la Conférence a approuvé les Lignes directrices à sa réunion de février 2006.

2. Le présent document porte le nom de «Lignes directrices provisoires», dans l'attente de son adoption officielle par la Conférence des statisticiens européens. Il sera soumis à la Conférence pour approbation à sa réunion plénière de 2006 et peut également être consulté sur l'Internet à l'adresse suivante: <http://www.unece.org/stats/documents/tfcm.htm.90>.

¹ Cette communication a été établie à l'invitation du secrétariat.

3. Les Lignes directrices constitueront un document dynamique auquel des mises à jour seront apportées périodiquement. On prévoit notamment que d'autres études de cas y seront intégrées. Les délégations sont invitées à formuler, au sujet de ce document, des observations qui, pour pouvoir être prises en considération par la Conférence à sa réunion plénière de juin 2006, devront être adressées à tiina.luige@unece.org avant le 19 mai 2006.

II. INTRODUCTION

4. Autrefois, la question de la protection de la confidentialité se posait essentiellement en termes nationaux mais, compte tenu de la diffusion croissante des données via l'Internet, elle prend aussi désormais une dimension internationale. Une collaboration internationale assez intensive s'est établie, dans ce domaine, au sein de la communauté des chercheurs qui, parfois, sont très critiques à l'égard des diverses règles d'accès en vigueur dans les différents pays. En outre, les chercheurs se voient souvent refuser l'accès aux microdonnées d'autres pays par crainte que la protection de la confidentialité ne puisse être garantie. Néanmoins, les comparaisons internationales peuvent constituer un élément très important d'un projet de recherche. Cet aspect de la question intéresse non seulement la recherche fondamentale, mais aussi, entre autres, les organisations internationales qui souhaitent exploiter les microdonnées dans le cadre de leurs études et notamment de comparaisons entre pays.

5. Il convient donc de s'interroger sur la possibilité d'arrêter à l'échelon international certains principes communs aux fins de la diffusion des microdonnées. Cette question devrait être traitée en partant du point de vue exprimé d'un commun accord par les membres de la Conférence des statisticiens européens en 2003, à savoir que l'appui à la recherche représente une activité importante des services nationaux de statistique et que ceux-ci pourraient de manière générale faire plus d'efforts pour répondre à ces besoins. Dans cette optique, il faudrait, entre autres, assurer l'accès aux microdonnées, point sur lequel sont principalement centrés les présents principes et lignes directrices (bien que ces lignes directrices fassent principalement mention des services nationaux de statistique, dans de nombreux pays, notamment ceux qui sont dotés de systèmes décentralisés, on compte plusieurs producteurs de statistiques. L'expression «services nationaux de statistique» doit donc s'entendre comme englobant tous les producteurs de statistiques officielles.)

6. Les lignes directrices répondent à deux objectifs essentiels:

a) Garantir une plus grande uniformité des démarches suivies par les pays tout en favorisant un meilleur accès de la communauté des chercheurs aux microdonnées; et

b) Permettre aux pays, par le biais de ces lignes directrices et des études de cas établies à l'appui, d'améliorer leurs dispositifs en matière d'accès aux microdonnées.

7. Le terme «microdonnées» est utilisé dans l'ensemble du document. Il peut s'agir de données concernant un individu, un ménage, une entreprise ou toute autre entité, et de données directement recueillies par le service national de statistique ou obtenues auprès d'autres sources, notamment de sources administratives.

8. Les lignes directrices reconnaissent que les modalités précises d'accès aux microdonnées varient d'un pays à l'autre. Elles dépendent de facteurs tels que la législation, l'attitude du public

et la capacité d'apporter un soutien à la communauté des chercheurs. Par exemple, le dispositif mis en place par un service de statistique très développé différera sensiblement de celui établi dans un service de statistique moins avancé. Il ne faut pas s'attendre à ce que tous les pays adoptent exactement les mêmes dispositions, même s'il y a lieu d'espérer que les lignes directrices permettront une démarche plus uniforme.

9. Il ne faudrait pas non plus perdre de vue que les pays ne sont pas tous égaux au départ. Certains d'entre eux, notamment ceux d'Europe orientale, n'étaient pas dotés précédemment d'une législation solide protégeant la confidentialité des données. Dans ce nombreux cas, des mesures sont prises pour remédier à cette lacune mais le changement d'attitude qui doit accompagner la modification des législations pourrait prendre plus de temps.

10. Plusieurs pays ont établi des dispositions législatives pertinentes. En outre, l'Union européenne (UE) a adopté une législation sur la confidentialité énonçant plusieurs principes et règles qui seront déjà appliqués par bon nombre de pays de la CEE, notamment des membres de l'Union européenne. On constate qu'il n'est pas facile de modifier la législation en vigueur et que des modifications des lignes directrices existantes appellent une collaboration avec toute une série de parties prenantes. Toutefois, l'occasion d'introduire des changements se présente de temps à autre et les lignes directrices peuvent être utiles pour déterminer les modifications qu'il convient d'apporter. En fait, dans certains pays, ces lignes directrices pourraient servir de déclencheur en ce sens qu'elles lanceraient le débat et conduiraient à arrêter des modifications.

11. Toute question portant sur les lignes directrices devrait être adressée par courrier électronique à la Division de statistique de la CEE à l'adresse suivante:
confidentiality@unece.org.

III. POURQUOI LES SERVICES NATIONAUX DE STATISTIQUE DEVRAIENT-ILS APPORTER LEUR APPUI À LA COMMUNAUTÉ DES CHERCHEURS?

12. Dans la plupart des pays, les statistiques officielles sont réunies non seulement à l'usage des pouvoirs publics mais aussi au profit de la collectivité. C'est notamment le cas dans les démocraties où les statistiques officielles peuvent servir à évaluer l'efficacité des politiques et programmes du gouvernement: elles constituent un reflet de la société.

13. Pour citer un livre blanc de 1993 sur la transparence des affaires publiques au Royaume-Uni:

«Des statistiques officielles sont recueillies par les pouvoirs publics dans le but d'étayer le débat, la prise de décisions et la recherche tant au sein de l'administration que de la collectivité en général.

Elles offrent une vue objective des modifications qui interviennent dans la vie nationale et permettent des comparaisons dans le temps et d'un lieu à l'autre.

Un accès libre aux statistiques officielles procure au citoyen davantage qu'une simple image de la société. Il lui permet de se faire une idée des travaux et de l'efficacité du gouvernement lui-même, en montrant l'ampleur de l'activité gouvernementale dans

chaque domaine de la politique publique et en permettant d'évaluer l'impact des politiques et de l'action des pouvoirs publics.»

14. La communauté des chercheurs joue un rôle particulièrement important dans le cadre des efforts qui visent à stimuler l'analyse des politiques et encourager le débat sur ces politiques, ainsi qu'à évaluer l'efficacité des programmes publics. Elle a besoin dans ce contexte d'avoir accès à des données statistiques de valeur pour que ses analyses soient efficaces. Si les chercheurs ne peuvent obtenir les données statistiques officielles qui leur sont nécessaires, ils cherchent généralement à recueillir leurs propres données. Outre les frais supplémentaires qu'elles occasionnent à la fois à ceux qui recueillent les données et aux répondants, ces collectes ne procurent souvent que des données médiocres.

15. En assurant aux chercheurs un accès aux microdonnées, on peut aussi mieux rentabiliser la collecte des statistiques officielles et avoir un avis précieux sur la qualité des données et sur la manière dont on pourrait améliorer les enquêtes statistiques ou en élargir la portée.

16. Qu'entend-on par communauté des chercheurs? Elle englobe évidemment les personnes qui travaillent dans des instituts de recherche universitaires, de même que les chercheurs appartenant à des organisations non gouvernementales et organisations internationales. En outre, certains chercheurs qui ont besoin d'avoir accès aux microdonnées oeuvrent au sein d'organismes et d'institutions financés par l'État. Aux fins des présentes lignes directrices, tous ces chercheurs sont pris en compte dans l'appellation «communauté des chercheurs». Toutefois, comme il ressort de ces lignes directrices, les questions pertinentes peuvent varier quelque peu selon l'élément de la communauté des chercheurs, que l'on considère.

17. Les sections ci-après tentent de regrouper les points de vue des services nationaux de statistique, d'une part, et ceux de la communauté des chercheurs, d'autre part, dans le but de trouver des modalités qui satisfassent largement les besoins des deux groupes. Cette question est examinée plus en détail au chapitre VII.

A. La perspective du service national de statistiques

18. Les services nationaux de statistique doivent conserver la confiance des personnes interrogées s'ils veulent que celles-ci continuent de collaborer à leurs collectes de données. La protection de la confidentialité est l'aspect clef de cette confiance. Si les répondants estiment ou ont l'impression qu'un service national de statistique ne va pas protéger la confidentialité des données qu'ils fournissent, ils seront moins enclins à collaborer aux enquêtes ou à communiquer des données précises. Un incident unique, particulièrement si les médias lui donnent un large écho, pourrait avoir un effet majeur sur le degré de coopération des répondants et, par conséquent, sur la qualité des statistiques officielles.

19. Cet aspect est prépondérant du point de vue des services nationaux de statistique, mais d'autres problèmes entrent en jeu. Un élément clef consiste à déterminer si ces services sont habilités à assurer des chercheurs aux microdonnées, soit par un instrument juridique, soit par une autorisation d'un autre type.

20. Certains services nationaux de statistique craignent que la qualité de leurs micro-données ne soit pas suffisante pour qu'elles puissent faire l'objet d'une plus large diffusion. Si elle peut

être suffisamment bonne pour permettre l'élaboration d'agrégats statistiques, la qualité de ces microdonnées peut en effet ne pas répondre aux exigences d'une analyse très fine. Dans certains cas, des ajustements sont apportés aux statistiques agrégées, au stade de l'édition du produit, sans que les microdonnées soient modifiées. Par conséquent, il peut y avoir des incohérences entre les résultats de la recherche fondée sur les microdonnées et les données agrégées publiées.

21. Les services nationaux de statistique peuvent également être préoccupés par la question des coûts, notamment ceux qu'impliquent la création et la documentation de fichiers de microdonnées mais aussi les frais associés à la mise en place d'instruments d'accès et de protection, ainsi qu'au soutien et à l'autorisation d'enquêtes réalisées par la communauté des chercheurs; en effet, les utilisateurs novices de leurs fichiers de données ont besoin d'aide pour s'y retrouver dans les structures de fichier complexes et les définitions de variables. Bien que les coûts en question soient à la charge des services nationaux de statistique, ceux-ci ne disposent généralement pas d'une enveloppe budgétaire supplémentaire pour financer les travaux supplémentaires à entreprendre dans ce contexte. Quant aux chercheurs, ils n'ont généralement pas les moyens d'assumer une substantielle de ces coûts.

22. Par ailleurs, les services nationaux de statistique sont de plus en plus conscients de l'importance que revêt l'appui à la communauté des chercheurs, et reconnaissent sans cesse davantage la valeur supplémentaire que l'exploitation efficace de leurs données aux fins de la recherche apporte à leurs efforts de collecte et de traitement des données. Plus précisément, il va de l'intérêt général que les connaissances, dégagées des données, puissent être mises à la disposition des décideurs et du public. En outre, une plus large exploitation des données par le secteur de la recherche peut assurer une protection supplémentaire contre la réduction des budgets des programmes statistiques.

B. La perspective de la communauté des chercheurs

23. Du point de vue de la communauté des chercheurs, le soutien de la recherche faisant appel aux microdonnées devrait représenter un élément important de n'importe quel système statistique officiel. Les avantages qui en découlent sont notamment les suivants:

a) Les microdonnées permettent aux décideurs de poser et d'analyser des questions complexes. En économie, par exemple, l'analyse des agrégats statistiques ne procure pas de vue suffisamment précise du fonctionnement de l'économie pour permettre l'analyse des composantes de la croissance de la productivité;

b) En ayant accès aux microdonnées, les analystes peuvent calculer des effets marginaux plutôt que de simples moyennes. Grâce aux microdonnées, ils procèdent par exemple à des régressions à plusieurs variables permettant d'isoler l'impact marginal de certaines d'entre elles;

c) De manière générale, un large accès aux microdonnées permet de reproduire des recherches importantes;

d) L'accès aux microdonnées à des fins de recherche et les réactions qu'il permet d'obtenir peuvent favoriser l'amélioration de la qualité des données. Par exemple, le Bureau of the Census (bureau des recensements) des États-Unis a défini avec précision les informations

qu'il souhaiterait avoir de la part des chercheurs pour pouvoir améliorer la qualité de ses enquêtes;

e) L'accès aux microdonnées élargit la gamme des produits élaborés à partir des statistiques recueillies et, par conséquent, la rentabilité globale des collectes de données.

24. Par ailleurs, le manque d'accès aux microdonnées peut conduire les chercheurs à mettre en place et réaliser leurs propres collectes de statistiques, ce qui vient alourdir la charge imposée à la collectivité. En dehors des coûts qu'implique ce genre de collecte (à la fois pour celui qui les organise et pour les répondants), les données réunies sont généralement de moindre valeur et s'appuient sur de plus petits échantillons que lorsqu'il s'agit d'enquêtes officielles. L'accès à une source de données reconnue et faisant autorité, de même que de haute qualité, présente des avantages pour toute forme d'analyse, par rapport à l'utilisation par les chercheurs de différents ensembles de données pour analyser des sujets particuliers.

25. Les chercheurs soulignent qu'ils ne se soucient pas d'identifier les individus et force est de constater que les faits le confirment. Cela étant, ils estiment que les services nationaux de statistique font généralement montre de trop de prudence quant à l'accès qu'ils offrent à leurs microdonnées.

26. Lors d'un atelier consacré en 2003 à l'étude de la confidentialité et organisé sous l'égide de la National Science Foundation des États-Unis, Peter Madsen a parlé de «paradoxe en matière de protection de la confidentialité». Il a fait valoir que le zèle déployé pour sécuriser entièrement les données dans le contexte de la recherche aboutit paradoxalement à un moindre avantage pour la société au lieu de le maximiser. Il a mis en avant qu'en prenant aussi en compte le concept d'utilité, on pourrait obtenir des résultats différents:

«Peut-être que si l'on ajoute ce concept d'utilité, les personnes interrogées prendront conscience que si elles ont sans nul doute le droit à la protection de leur vie privée, elles ont également le devoir de communiquer des informations si celles-ci servent l'intérêt général.»

D'aucuns évoquent un «déficit de protection de la confidentialité» du fait que la diffusion des microdonnées soulève certaines questions en la matière. Il s'agit ensuite de déterminer si les avantages d'un projet de recherche l'emportent sur toute carence dans la protection de la confidentialité.

27. La communauté des chercheurs juge également important d'étudier des méthodes améliorées de protection de la confidentialité susceptibles d'accroître l'utilité des données de base. Les services nationaux de statistique reconnaissent eux aussi l'importance de telles études. Toutefois, ces études ne vont sans doute apporter qu'une réponse partielle au souci d'améliorer l'accès aux microdonnées à des fins de recherche, et les chercheurs se sentiraient frustrés si les services nationaux de statistique s'appuyaient uniquement sur de meilleures méthodes statistiques pour assurer la protection de la confidentialité.

IV. PRINCIPES DE BASE

28. Le sixième des Principes fondamentaux de la statistique officielle de l'ONU est explicite sur la question de la confidentialité:

«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.».

N'importe quel principe d'accès aux microdonnées doit être conforme à ce Principe fondamental.

29. Les principes ci-après devraient régir la gestion de la confidentialité des microdonnées. Chacun d'eux est examiné plus en détail dans les paragraphes qui suivent:

a) Principe 1: On peut exploiter les microdonnées réunies dans le cadre de la statistique officielle, aux fins de l'analyse statistique, en vue d'étayer des recherches pour autant que la confidentialité de ces données soit protégée;

b) Principe 2: Les microdonnées ne devraient être communiquées qu'à des fins statistiques;

c) Principe 3: La communication de microdonnées devrait s'effectuer en accord avec les dispositions juridiques et autres nécessaires pour garantir la confidentialité des microdonnées publiées;

d) Principe 4: Il faudrait assurer la transparence des modalités d'accès des chercheurs aux microdonnées, ainsi que des utilisations et utilisateurs de microdonnées, et les rendre publics.

30. La communication des microdonnées aux chercheurs n'est pas incompatible avec le sixième Principe fondamental de l'ONU tant qu'il demeure impossible d'identifier les données se rapportant à un individu. Le Principe 1 ci-dessus ne constitue pas une obligation de diffuser les microdonnées. C'est au service national de statistique qu'il appartient de décider s'il convient ou non de fournir les microdonnées. D'autres considérations (la qualité des microdonnées, par exemple) peuvent rendre inopportun l'accès aux microdonnées. Il se peut aussi qu'il soit inapproprié de fournir des microdonnées à des personnes ou institutions bien précises.

31. Dans le cas du Principe 2 susmentionné, il faut faire la distinction entre les exploitations statistiques ou analytiques des microdonnées et leur utilisation à des fins administratives. Dans le cas des exploitations statistiques ou analytiques, l'objectif consiste à établir des statistiques se rapportant à un groupe (qu'il s'agisse de personnes physiques ou morales). Pour ce qui est de l'utilisation administrative, le but recherché est d'obtenir des informations sur une personne physique ou morale en vue de prendre une décision susceptible de profiter ou de nuire à un particulier. Par exemple, certaines demandes de données peuvent être légales (une décision judiciaire, notamment) mais incompatible avec le Principe 2. Pour promouvoir la confiance du public dans le système statistique officiel, ces demandes doivent être systématiquement rejetées. Si l'utilisation escomptée des microdonnées est incompatible avec les fins statistiques ou

analytiques, il ne faudrait pas donner accès aux microdonnées. Les comités d'éthique ou tout système analogue peuvent prêter leur concours dans les situations où une certaine latitude doit s'exercer quant à l'accessibilité des microdonnées.

32. Les chercheurs ont accès aux microdonnées à des fins de recherche mais pour étayer cette recherche, ils peuvent avoir besoin d'élaborer des agrégats statistiques de nature diverse, d'établir des distributions statistiques, d'ajuster des modèles statistiques, ou d'analyser les différences statistiques entre les sous-populations. Les utilisations de cette nature seraient également conformes aux fins statistiques. Pour autant que les microdonnées soient exploitées ainsi, on peut également considérer que ces utilisations répondent aux exigences de la recherche.

33. En ce qui concerne le Principe 3, il faudrait mettre en place des dispositions juridiques visant à protéger la confidentialité avant de publier quelque microdonnée que ce soit. Toutefois, les dispositions juridiques doivent être complétées par des mesures administratives et techniques tendant à réglementer l'accès aux microdonnées et à faire en sorte que les données personnelles ne puissent être divulguées. L'existence et la visibilité de ce genre de modalités (qu'elles soient consacrées par la loi ou des règlements supplémentaires, des arrêtés, etc.) sont indispensables pour donner au public une plus grande confiance dans le bon usage qui sera fait des microdonnées. De toute évidence, la meilleure solution consiste à instaurer des dispositions juridiques mais dans certains pays cela ne se peut. Il faudrait donc alors instituer une autre forme quelconque de dispositions administratives. Il faudrait également que les autorités responsables des questions de confidentialité, dans les pays où il en existe, donnent leur accord aux dispositions juridiques (ou autres) avant que celles-ci ne soient consacrées par la loi. En l'absence d'autorité de cette nature, il peut y avoir des ONG qui exercent une fonction de «surveillance» dans le domaine de la confidentialité. Il serait judicieux d'obtenir qu'elles souscrivent à toute forme de disposition juridique ou autre mise en place, ou tout au moins de répondre à n'importe quelle inquiétude sérieuse qu'elles pourraient avoir.

34. Dans certains pays, il n'existe pas de législation habilitante. Il faudrait au minimum que la diffusion des microdonnées s'appuie sur une forme quelconque d'autorité. Cependant, l'existence d'une législation habilitante est la démarche préférable.

35. Le Principe 4 est important pour rassurer le public quant au fait que les microdonnées sont exploitées à bon escient et pour montrer que les décisions touchant la diffusion des microdonnées sont prises sur une base objective. Il appartient au service national de statistique de déterminer si les microdonnées peuvent être divulguées, selon quelles modalités et à quel utilisateur. Néanmoins, il faudrait assurer la transparence de ses décisions. Le site Web du service national de statistique constitue un instrument efficace pour garantir le respect des règles établies mais aussi pour fournir des informations sur les moyens d'avoir accès aux rapports d'études fondées sur les microdonnées publiées.

V. FONDEMENTS JURIDIQUES

36. Il est capital que la diffusion des microdonnées s'appuie sur des fondements juridiques, ainsi qu'il ressort du Principe 3 (voir le chapitre IV du présent document), et ce pour plusieurs raisons:

a) Pour que le public ait confiance dans les dispositions établies, en d'autres termes qu'il sache que des contraintes juridiques déterminent ce qui est ou non autorisé;

- b) Pour qu'une compréhension mutuelle s'instaure entre les services nationaux de statistique et les chercheurs en ce qui concerne ces dispositions;
- c) Pour qu'une plus grande cohérence soit assurée dans le traitement des projets de recherche; et
- d) Pour que soient mis en place les fondements d'un système visant à sanctionner les infractions.

37. Il n'est pas indispensable que les dispositions en question soient énoncées dans une loi. Les détails de la protection des microdonnées peuvent plus aisément être précisés par le biais de règlements, d'arrêtés, etc., qui n'en ont pas moins un effet juridique. En l'absence de dispositions juridiques, l'une ou l'autre forme d'autorisation est essentielle. La réputation du service national de statistique sera compromise s'il n'existe pas une forme quelconque d'autorité s'exerçant sur la divulgation des microdonnées, même anonymisées.

38. Il importe que la législation (ou l'autorisation) prenne en compte les aspects suivants:

- a) Ce qui est ou non autorisé et à quelles fins;
- b) Les conditions de la divulgation; et
- c) Les conséquences d'un non-respect de ces conditions.

39. Les études de cas 1 et 2 font état des dispositions juridiques régissant la fourniture de microdonnées en Australie et en Finlande, respectivement.

VI. MÉTHODES D'APPUI À LA COMMUNAUTÉ DES CHERCHEURS

40. Il y a diverses méthodes qu'un service national de statistique peut mettre en œuvre pour appuyer les travaux de recherche et qui sont résumées ci-après. On trouvera des commentaires plus détaillés dans les paragraphes qui suivent. Ces différentes méthodes sont illustrées avec plus de précision par des études de cas.

- a) Produits statistiques utilisables en dehors du service national de statistique:

Flux de diffusion	Notes
Tableaux statistiques et cubes de données	Il peut s'agir à la fois de tableaux types et de tableaux spéciaux (ou, sinon, d'analyses spéciales) établis à la demande du chercheur. Certains services publient à présent des matrices très détaillées, appelées «cubes de données», que les chercheurs peuvent manipuler en fonction de leurs propres besoins. Toutefois, si ces matrices sont très détaillées, le risque qu'elles présentent du point de vue de la confidentialité peut être du même ordre que celui qui se rattache aux microdonnées.

Fichiers de microdonnées anonymisés	Il s'agit de fichiers de microdonnées destinés à l'usage du grand public en dehors du service national de statistique. Ils ont été anonymisés et sont souvent divulgués sur un support tel qu'un CD-ROM, parfois par le biais d'une archive de données (Note: le terme «anonymisé» signifie que les noms et adresses sont non seulement éliminés mais aussi que d'autres mesures sont prises pour rendre très improbable l'identification des individus). Le degré de protection de la confidentialité qu'offrent les fichiers à usage public, même lorsqu'ils sont mis en correspondance avec d'autres fichiers de données, devrait être tel qu'aucun engagement préalable ne s'impose pour en assurer la confidentialité. Dans bon nombre de pays, les fichiers à usage public constituent un moyen courant de permettre aux chercheurs d'avoir accès aux données.
– Fichiers à usage public	

Fichiers de microdonnées anonymisés	Les fichiers sous licence sont aussi anonymisés mais se distinguent des fichiers à usage public en ce sens que leur emploi peut être réservé à des chercheurs autorisés et qu'un engagement ou un contrat est signé avant que les fichiers leur soient communiqués. Même s'il est annoncé que ces fichiers sont librement mis à la disposition du public, ils ne sont pas diffusés avant qu'un engagement ou un contrat soit signé par le chercheur. Après avoir été anonymisés, il arrive qu'ils contiennent encore des données potentiellement identifiables si on les rapproche d'autres fichiers de données; c'est l'une des raisons pour lesquelles un engagement ou un contrat est nécessaire à titre préventif.
– Fichiers sous licence	

b) Un guichet par le biais duquel les chercheurs peuvent soumettre leur demande de données

Service	Notes
Moyens d'accès à distance	Des dispositions sont actuellement prises dans bon nombre de pays pour autoriser les chercheurs à élaborer des produits statistiques à partir de fichiers de microdonnées par le biais de réseaux informatiques, sans que les chercheurs «voient» effectivement les microdonnées. Grâce aux contrôles supplémentaires que les moyens d'accès à distance permettent d'exercer et au fait que les microdonnées ne quittent pas effectivement le service national de statistique, on peut assurer l'accès à des microdonnées plus détaillées.

- c) Dispositions autorisant les chercheurs à travailler dans les locaux du service national de statistique

Service	Notes
Laboratoires de données	Accès sur place à des microdonnées plus aisément identifiables, généralement via un circuit de vérification strict et sous la supervision du service national de statistique. L'accès à des données plus détaillées crée certaines complications pour le chercheur dans la mesure où il est tenu de travailler dans les locaux ou une antenne du service national de statistique.

A. Tableaux statistiques et cubes de données

41. Les tableaux statistiques restent le moyen le plus économique de satisfaire bon nombre de besoins en matière de recherche. Il ne faudrait donc pas sous-estimer leur importance. Avec l'apparition des cubes de données (des tableaux pluridimensionnels très détaillés), l'utilité des tableaux statistiques aux fins de la recherche s'est accrue car les chercheurs peuvent manipuler ces cubes de données en fonction de leurs propres besoins.

42. Statistics Netherlands est l'un des premiers organismes à avoir adopté les cubes de données et l'étude de cas 3 illustre la manière dont il s'en sert dans le cadre de sa stratégie de diffusion.

43. Des problèmes de confidentialité continuent de se poser en ce qui concerne les tableaux statistiques et cubes de données. Par exemple, la plupart des dispositions législatives en matière de statistique précisent qu'il n'est pas possible de divulguer des données identifiables sous la forme de tableaux statistiques. Toutefois, les données sont «confidentialisées» avant leur publication. Il existe des systèmes logiciels qui permettent de confidentialiser les tableaux statistiques et les méthodes ne cessent de s'améliorer. On les désigne souvent sous le nom de méthodes de non-divulgaration.

B. Fichiers de microdonnées anonymisés – Fichiers à usage public

44. Les chercheurs considèrent qu'ils retirent un service très utile de ces fichiers. Toutefois, compte tenu des possibilités accrues de mise en correspondance des données, on pourrait évoluer vers une réduction du volume des données disponibles sous la forme de fichiers de microdonnées anonymisés et s'appuyer davantage sur des services tels que les moyens d'accès à distance et les laboratoires de données pour permettre aux chercheurs d'avoir accès à l'information. Une autre solution consisterait à se fier davantage au fait que les chercheurs honoreront l'engagement qu'ils ont pris ou le contrat qu'ils ont signé en ne cherchant pas à identifier les individus ou entités auxquels se rapportent les données. Ce genre d'engagement fait en général partie intégrante de la publication de fichiers de microdonnées anonymisés (voir la section suivante).

45. Bien que les services nationaux de statistique garantissent généralement un accès égal à tous les utilisateurs de leurs statistiques, ce principe ne vaut pas nécessairement pour les microdonnées. Une attitude différente peut être adoptée à l'égard d'utilisateurs qui ne sont pas

pleinement accrédités en qualité de chercheurs ou qui ont accès à des bases de données par le biais desquelles ils pourraient facilement apparier les fichiers de microdonnées anonymisés.

46. Font exception à cet égard les fichiers à usage public auxquels on assure délibérément un large accès. Les chercheurs ont mis l'accent sur l'importance de ces fichiers. Ils sont très appréciés dans les pays où il en existe et où ils sont largement exploités. Toutefois, une personne qui y est disposée pourrait n'avoir aucune difficulté à repérer et identifier certains individus par le biais d'un appariement avec d'autres bases de données, notamment dans les pays peu peuplés et dans ceux qui sont dotés de registres de la population. Avant de divulguer les données de fichiers à usage public, il convient d'étudier de près les conditions dans lesquelles ceux-ci sont diffusés pour mieux gérer les risques de violation de la confidentialité. Par exemple, un engagement juridiquement contraignant peut constituer l'une des conditions d'accès. Il devrait être possible de mettre en place ce genre de disposition dans les cas où un engagement préalable doit être signé, même lorsque l'accès au fichier à usage public se fait par le biais de l'Internet. De manière générale, le degré de risque sera beaucoup plus important dans les pays à plus faible population. Par conséquent, les chercheurs ne devraient pas s'attendre à ce que tous les pays divulguent leurs fichiers à usage public.

47. L'étude de cas 4 fait état des dispositions relatives à la publication de fichiers à usage public aux États-Unis. Il est intéressant de noter le rôle que les archives de données sociales jouent dans la gestion de l'accès des différents chercheurs aux fichiers à usage public.

48. Il existe une abondante documentation sur les méthodes permettant d'anonymiser les fichiers de microdonnées. On trouve un bon résumé en la matière dans Willenburg, L. & de Waal, T (2001), Elements of Statistical Disclosure Control. Le progiciel μ -ARGUS sert à protéger les microdonnées de la divulgation. Il propose plusieurs techniques.

C. Fichiers de microdonnées anonymisés – Fichiers sous licence

49. Il s'agit d'un système qui autorise certains utilisateurs à se servir de fichiers de microdonnées ou leur délivre la licence nécessaire à cet effet. La licence est assortie de certaines conditions, qui peuvent être énoncées dans un engagement ou un contrat signé par le chercheur ou l'organisme dont il relève. Les conditions peuvent varier d'un pays à l'autre, voire d'un chercheur à l'autre, selon le projet de recherche et, éventuellement, l'entité à laquelle le chercheur est rattaché. Il peut s'avérer souhaitable de signer un engagement ou un contrat même dans les cas où l'on encourage la mise à disposition du public de ces fichiers de microdonnées.

50. Les conditions prévues peuvent notamment être les suivantes:

- a) Un accord selon lequel le chercheur s'engage à respecter les conditions de publication;
- b) Un engagement à ne pas chercher à identifier des individus ou des organisations;
- c) La certification que l'information ne servira qu'à des fins statistiques ou pour la recherche;

d) L'engagement de ne pas communiquer les microdonnées à d'autres personnes; la promesse de rendre les microdonnées au service national de statistique dès l'achèvement du projet de recherche; et

e) La renonciation à appairer les statistiques avec d'autres bases de données sans autorisation.

51. Il est de bonne règle que ce genre d'engagement s'appuie sur un certain fondement juridique, par exemple qu'il soit incorporé dans une législation habilitante. Ainsi, des mesures juridiques pourraient être prises à l'égard des auteurs d'infractions. Cela n'empêche pas l'adoption d'autres mesures s'agissant des infractions, telles que le refus de fournir d'autres services au chercheur concerné et/ou éventuellement à l'organisme dont il relève. Ces questions sont traitées au chapitre VIII du présent document.

52. Il devrait être possible de diffuser plus de données par le biais de fichiers sous licence que par le truchement de fichiers à usage public, dans la mesure où l'on peut aussi se reposer sur l'engagement pris par le chercheur d'assurer la protection de la confidentialité des données. Autrement dit, ce moyen de diffusion est plus fiable dans les cas où certaines données sont potentiellement identifiables par appariement avec d'autres fichiers.

53. Quelques pays diffusent, à l'extérieur, des fichiers de microdonnées contenant des données confidentielles. Certes, cette diffusion s'effectue en vertu d'accords de licence stricts qui précisent les conditions auxquelles les données peuvent être exploitées. Néanmoins, ce genre d'arrangement ne correspond pas à une bonne pratique à moins qu'il ne soit consacré par la législation. Mais même dans ce cas le risque existe que le public considère cette utilisation comme inappropriée et il pourrait en résulter une baisse de la participation aux enquêtes.

54. Les études de cas 5, 6 et 7 décrivent les modalités de publication de fichiers de microdonnées sous licence en Australie, aux Pays-Bas et en Suède, respectivement.

D. Moyens d'accès à distance

55. Ces moyens de diffusion revêtent une importance croissante mais la manière dont ils sont mis en œuvre varie sensiblement d'un pays à l'autre. Leur caractéristique principale réside dans le fait que les chercheurs n'ont pas accès aux microdonnées proprement dites mais peuvent faire exécuter à distance les tâches nécessitant ces microdonnées. Fréquemment, un arrangement contractuel a été conclu à cette fin entre le service national de statistique et le chercheur ou l'organisme auquel ce dernier est rattaché.

56. À titre d'exemple, Statistique Canada fournit aux chercheurs des fichiers de microdonnées fictifs et les autorise à soumettre des demandes d'exploitation en relation avec le fichier complet via le réseau informatique. Il traite les demandes hors ligne et expédie les résultats par le biais de réseaux informatiques après avoir effectué un contrôle de la confidentialité des données. Bien qu'il existe des dispositions semblables au Bureau australien de statistique, on note aussi des différences importantes. Les fichiers de microdonnées sont confidentialisés pour empêcher toute identification spontanée avant d'être rendus accessibles à distance. Toutefois, les exploitations expérimentales sont autorisées en relation avec des fichiers accessibles à distance et un petit nombre d'enregistrements unitaires non identifiables peuvent être téléchargés dans le but

d'étudier les valeurs aberrantes, etc. Ce produit est contrôlé avant d'être expédié aux chercheurs. Le système assure actuellement une exploitation par lots mais la mise au point d'une version interactive est en cours. Les modalités établies à Statistics Denmark sont encore différentes. Il s'agit d'un système en ligne grâce auquel les chercheurs peuvent effectuer des analyses en regard du fichier de microdonnées complet. Les modalités sont conçues de telle manière que le téléchargement des microdonnées elles-mêmes est impossible. Pour mieux gérer les risques, elles font plutôt appel aux accords conclus par les institutions et aux sanctions (notamment le refus de tout accès futur) si les règles ne sont pas respectées.

57. Il existe deux types fondamentaux de système d'accès à distance:

a) L'exécution à distance, qui permet à un chercheur de soumettre un programme et de recevoir ultérieurement le produit par courriel;

b) Les services à distance, grâce auxquels le chercheur exécute lui-même l'analyse et voit immédiatement la réponse s'afficher sur l'écran.

Bon nombre de pays ont opté pour la formule a) mais, mis à part le système danois, les modalités comme celles indiquées sous b) en sont encore au stade de la mise au point. L'acceptabilité des différents systèmes risque de varier d'un pays à l'autre.

58. Bien qu'ils ne soient jusqu'ici en place que dans quelques pays, et quoique les modèles et démarches varient, comme illustré ci-dessus, l'expérience des moyens d'accès à distance est, à ce jour, généralement positive.

59. Du point de vue des coûts, les moyens d'accès à distance sont préférables aux laboratoires de données (voir ci-après) puisque l'accès sous contrôle dans un tel système exige moins de main-d'œuvre que l'utilisation supervisée dans le cadre des laboratoires de données.

60. Si un tel système n'élimine pas entièrement le risque d'identification, un accord sous l'une ou l'autre forme doit néanmoins être signé par les chercheurs pour faire en sorte qu'ils soient pleinement conscients de leurs obligations. Il est de bonne règle de réserver l'accès aux données aux chercheurs qui ont signé un accord quelconque énonçant les conditions d'accès. La formation revêt également son importance, de même que le suivi et le contrôle périodiques de l'utilisation de ces moyens.

61. Les études de cas 8, 9 et 10 décrivent les moyens d'accès à distance mis en place au Canada, en Australie et au Danemark respectivement.

E. Laboratoires de données

62. Ils existent depuis de nombreuses années dans quelques services nationaux de statistique et assurent avec efficacité la protection contre le risque d'identification tout en permettant aux chercheurs d'avoir notamment accès à des ensembles de données pour lesquels il n'est pas possible de diffuser un fichier de microdonnées confidentialisé. Ils vont néanmoins de pair avec des conditions d'accès pour garantir un degré de protection suffisant. La principale critique formulée à l'égard des laboratoires de données concerne leur manque de commodité pour le chercheur qui se voit, par exemple, parfois contraint à se servir de logiciels d'analyses de

données qu'il connaît mal. Leur gestion est également coûteuse pour les services nationaux de statistique par rapport à d'autres solutions.

63. Certains services nationaux de statistique ont construit des nouveaux locaux pour héberger les laboratoires de données dans des endroits plus accessibles pour les chercheurs (parfois connus sous le nom de «Centres de recherche de données»), mais cette solution peut également s'avérer coûteuse à moins que des fonds soient spécifiquement alloués à cet effet au service national de statistique.

64. Quelles sont les conditions essentielles d'accès aux microdonnées par le biais de laboratoires de données? Celles-ci pourraient inclure: a) la fourniture d'une documentation attestant de l'intérêt général qui découlera de la recherche, b) une description de l'accessibilité des résultats au public, c) des preuves de la bonne foi des chercheurs, d) un engagement juridiquement contraignant, et e) des exigences concernant la supervision exercée par le service national de statistique.

65. Les études de cas 11, 12, 13 et 15 décrivent les modalités mises en place en matière de laboratoires des données au Canada, aux États-Unis, aux Pays-Bas, en Nouvelle-Zélande et en Italie, respectivement.

F. Recrutement d'un chercheur en qualité de fonctionnaire temporaire d'un service national de statistique

66. Un autre moyen de veiller à ce que les chercheurs aient accès aux microdonnées consiste à les recruter comme agents temporaires d'un service national de statistique et de les soumettre aux mêmes règles relatives à la confidentialité que le personnel du service en question. En fait, on devrait alors considérer leurs recherches comme une contribution au programme de travail du service national de statistique plutôt que de les envisager sous l'angle de l'accès aux données qui leur est ainsi proposé. Il ne faudrait avoir recours à cette formule que si l'activité du chercheur représente un véritable apport au travail du service national de statistique; sinon on pourrait n'y voir qu'un simulacre. Si un arrangement de cette nature n'était que de la comédie et si cela se savait, le service national de statistique perdrait de sa crédibilité.

67. Le chercheur peut être associé aux travaux du service national de statistique à l'initiative de celui-ci, s'il considère que cette personne est susceptible de le faire bénéficier de compétences particulières et d'accroître l'utilité de l'ensemble de données. Par ailleurs, l'initiative peut venir du chercheur. Mais, dans ce cas, le service national de statistique doit jauger et accepter la proposition selon ses mérites et l'incorporer aux activités de son programme de travail. Il est plus facile de prouver que les chercheurs apportent leur concours au service national de statistique si les résultats de leurs travaux sont publiés par le service (même si c'est sous une appellation quelque peu différente de celle de ses produits usuels).

G. Données se rapportant aux entreprises

68. Les données des entreprises, notamment celles des exploitations agricoles, soulèvent quelques problèmes particuliers. Les entreprises, et notamment les plus grandes d'entre elles, sont plus facilement identifiables qu'un ménage ou un individu, surtout de manière spontanée, parce que la distribution de leurs caractéristiques est beaucoup plus asymétrique. Dans certains

pays, les bases de données sur les entreprises sont souvent plus accessibles, ce qui permet le rapprochement des données. En outre, il arrive que bon nombre de chercheurs fassent également office de consultants auprès des entreprises et, même en bonne et due forme, l'accès de ces chercheurs aux microdonnées des entreprises peut s'avérer incompatible avec leur mission de consultant (en effet, on ne pourra pas effacer de leur mémoire les connaissances qu'ils auront acquises durant leurs recherches). Par ailleurs, les pays peuvent être confrontés à des questions de compétitivité économique (voire de sécurité) du fait qu'ils échangent des données d'entreprises identifiables avec des chercheurs d'autres pays.

69. Pour ce qui est de l'accès des chercheurs aux données, la principale différence entre les données se rapportant aux ménages ou aux personnes, d'une part, et les données d'entreprises, d'autre part, tient au fait que les flux de diffusion assurant la plus grande protection concernent surtout les données des entreprises.

70. Quant aux moyens de diffusion:

a) Les tableaux statistiques restent un moyen intéressant de diffusion de données lorsqu'il s'agit d'entreprises; toutefois, comme ils présentent un risque d'identification plus élevé dans ce cas, ils ne contiennent généralement pas de données très détaillées;

b) Les fichiers de microdonnées anonymisés peuvent n'être pertinents que pour les plus petites entreprises. Dans certaines études, ce groupe intéresse particulièrement les chercheurs. Il n'en reste pas moins que quelques données (par exemple, des données financières) devront être «altérées» pour qu'elles ne puissent pas être mises en correspondance avec d'autres bases de données (par exemple, des données fiscales). Une autre solution consiste à présenter les données sous la forme de fourchettes de valeurs. En tout état de cause, les fichiers de microdonnées anonymisés n'auront probablement qu'une utilité limitée;

c) Pour des raisons analogues, les moyens d'accès à distance peuvent n'être utiles que dans le cas des fichiers de microdonnées concernant les plus petites entreprises. L'utilisation de ces systèmes permettra tout au moins aux services nationaux de statistique de contrôler le risque d'appariement afin qu'il ne soit pas forcément nécessaire d'«altérer» les données pour en protéger la confidentialité. Cependant, si les grandes entreprises sont prises en compte, il pourrait s'avérer difficile de confidentialiser les produits publiés même si les chercheurs ne peuvent pas avoir directement accès aux microdonnées;

d) Les laboratoires de données sont sans doute le mode le plus pertinent d'accès aux fichiers de microdonnées des entreprises. Ce genre de modalités d'accès existe, par exemple, à Statistics Netherlands.

71. Des microdonnées pourront sans doute être communiquées pour certaines études avec l'assentiment des entreprises concernées.

VII. GESTION DES DIFFÉRENCES DE POINTS DE VUE ENTRE LES SERVICES NATIONAUX DE STATISTIQUE ET LES CHERCHEURS

A. Comment pourrait-on concilier les perspectives différentes du service national de statistique et des chercheurs?

72. Ce sont les services nationaux de statistique qui pourront le plus efficacement concilier ces différentes perspectives en passant d'une stratégie d'évitement des risques à une stratégie de gestion des risques. Les moyens à mettre en œuvre à cette fin sont examinés plus en détail dans les paragraphes ci-après.

73. De toute évidence, des risques existent et il faut les gérer. L'expansion rapide des bases de données, contenant des informations sur des personnes identifiables, signifie qu'il est pratiquement impossible d'éliminer complètement le risque d'identification d'un nombre important de personnes même en supprimant leurs noms et adresses, surtout si la structure du ménage est indiquée dans les fichiers. Les bases de données en question sont pour beaucoup tenues par le secteur privé qui en contrôle généralement l'emploi de manière moins stricte que le secteur public. En outre, les progrès techniques facilitent le rapprochement des données, qu'il s'agisse d'une correspondance exacte ou de méthodes d'appariement statistique (susceptibles de conduire, exceptionnellement, à des correspondances exactes). En substance, l'évitement des risques signifie que les microdonnées identifiables ne doivent pas quitter les locaux du service national de statistique (on notera que les risques varient, entre autres, selon la taille du pays. Dans les petits pays de taille plus restreinte, le risque est relativement élevé parce qu'il y a plus de cas exceptionnels).

74. Néanmoins, l'accès aux microdonnées assuré par les services nationaux de statistique ne semble pas donner lieu à une controverse publique. On peut en déduire que le public accepte relativement bien les pratiques actuelles, bien qu'à notre connaissance, aucun pays n'ait organisé un débat ouvert sur la question. Toutefois, le souci général du public en matière de protection de la confidentialité donne à penser qu'il y a une limite à ce qu'il est prêt à accepter. Une controverse pourrait aisément être déclenchée (au-delà des frontières nationales) par un seul incident malheureux, ce qui risquerait de nuire à la participation aux collectes ultérieures de données statistiques.

75. La transparence est importante pour pallier les accusations d'excès de confidentialité. Par conséquent, il est souhaitable que les services nationaux de statistique fassent ouvertement savoir que l'une des utilisations précieuses des données tirées de certaines collectes consiste à permettre aux chercheurs de se servir de microdonnées confidentielles dans des conditions strictes et exclusivement à certaines fins. Cet aspect des choses doit être géré avec prudence, sinon les inconditionnels de la confidentialité risquent de retourner l'opinion publique. Il est essentiel que des personnes respectées et faisant autorité prêtent leur appui à cet égard.

B. Comment les services nationaux de statistique gèrent-ils les risques?

76. Quelques suggestions sont présentées ci-après:

a) Arrêter une série de principes qu'il faudrait suivre pour donner accès aux microdonnées (notamment ceux énoncés au chapitre IV du présent document);

b) Faire en sorte qu'on dispose d'une base juridique et éthique solide (ainsi que des instruments techniques et méthodologiques nécessaires) pour protéger la confidentialité tout en offrant un accès aux microdonnées. Cette base juridique et éthique nécessite que l'on mette soigneusement en balance l'intérêt pour le public d'une solide protection de la confidentialité, d'une part, et les avantages que lui apporte la recherche, d'autre part. La décision de savoir s'il convient ou non de permettre à un chercheur d'avoir accès aux données pourrait dépendre des mérites de son projet de recherche et de sa crédibilité, et il faudrait en tenir compte d'une façon ou d'une autre dans les fondements éthiques. On ne doit pas considérer l'accès comme étant automatique. Les services nationaux de statistique sont tenus de se conformer aux dispositions législatives ou autres en vigueur dans leur pays, quel que soit l'intérêt de la proposition;

c) Mettre en place une procédure indépendante pour évaluer en regard l'une de l'autre les deux catégories d'avantages susmentionnées. Il serait judicieux de créer un comité interne chargé de débattre de ces questions et d'adresser des recommandations au chef du service national de statistique. Des comités d'éthique peuvent également prêter leur concours dans des situations où une certaine latitude doit s'exercer pour déterminer s'il convient ou non de diffuser les données. Les arguments plaidant en faveur de l'intérêt général seront beaucoup plus frappants s'il est prévu de placer les résultats de la recherche dans le domaine public;

d) Assurer une transparence complète concernant les utilisations particulières des microdonnées pour dissiper tout doute quant à une éventuelle utilisation abusive;

e) Être prêt à offrir aux chercheurs une plus grande facilité d'accès par le biais des moyens d'accès à distance et des laboratoires de données car il peut s'avérer impossible de rendre les microdonnées destinées à la publication complètement non identifiables sans altérer sensiblement les données. Étudier d'autres possibilités de recours aux progrès technologiques pour améliorer l'accès aux microdonnées de telle sorte qu'une protection suffisante de la confidentialité soit assurée;

f) Transmettre une part de la responsabilité à la communauté des chercheurs. Faire en sorte qu'elle comprenne pleinement les raisons pour lesquelles les services nationaux de statistique sont tellement attachés à la protection de la confidentialité. S'assurer que les chercheurs soient conscients des conséquences d'infractions éventuelles pour eux-mêmes et l'organisme dont ils relèvent. Prévoir des sanctions éventuelles en cas d'infraction.

77. Le dernier point appelle certains commentaires. La culture et les valeurs des chercheurs sont très différentes de celles d'un service national de statistique. Les chercheurs considèrent souvent certains des «contrôles» inhérents au processus d'accès aux microdonnées, comme trop bureaucratiques. S'il n'y a aucun incident connu d'utilisation abusive par les chercheurs de leur accès aux microdonnées pour identifier volontairement des individus, il est arrivé que des chercheurs communiquent à des collègues sans y avoir été autorisés des microdonnées qui leur avaient été fournies pour leur usage exclusif, ou que des chercheurs mettent des microdonnées en correspondance avec d'autres données, sans autorisation préalable, dans le but d'élaborer des ensembles de données plus étoffés. Les chercheurs en question peuvent avoir l'impression qu'ils n'ont rien fait de mal car ils n'ont pas tenté d'identifier des individus. Toutefois, des incidents de cette nature, s'ils sont divulgués, peuvent saper la confiance du public. Les services nationaux de statistique et les chercheurs ont des cultures différentes et des conceptions différentes des risques

découlant de ce genre d'incidents. Il faut en tenir compte pour définir des procédures de diffusion des microdonnées.

78. Comment les services nationaux de statistique peuvent-ils transférer une partie des risques aux chercheurs? Les mesures envisageables à cet effet pourraient notamment être les suivantes:

- a) Leur demander de prouver leur légitimité en qualité de chercheurs, de démontrer l'intérêt de leurs recherches pour le public et de fournir des preuves que les microdonnées sont «adaptées à l'objectif recherché»;
- b) Faire en sorte que les chercheurs signent un engagement juridiquement contraignant prévoyant des sanctions analogues à celles qui s'appliquent aux fonctionnaires du service national de statistique s'ils enfreignent les règles en matière de confidentialité;
- c) Expliquer les raisons de la prudence dont font preuve les services nationaux de statistique. S'assurer que les chercheurs sont pleinement conscients de leurs obligations en les informant comme il se doit. Instaurer des procédures de suivi et de surveillance efficaces. Il pourrait s'avérer utile d'établir un code de conduite en collaboration avec la communauté des chercheurs;
- d) Lorsque des infractions sont commises, refuser au chercheur et éventuellement à l'organisme dont il relève de leur fournir tout service pendant un certain temps (par exemple jusqu'à ce que l'organisme ait pris les mesures disciplinaires qui s'imposent à l'égard du chercheur en faute). Entreprendre une action en justice s'il convient.

79. En réalité, un ensemble de mesures juridiques, administratives et techniques devra être mis en place pour gagner et conserver la confiance du public. En outre, la communauté des chercheurs doit accepter le fait qu'elle ne bénéficie pas d'un droit d'accès automatique. Un service national de statistique peut être habilité de manière générale à donner accès à ses données mais la décision d'autoriser ou non les chercheurs à consulter ces données devrait être dans chaque cas laissée à la discrétion du service en question. Le droit d'accès devrait s'accompagner de responsabilités. En particulier, les chercheurs devraient accepter de partager la responsabilité de préserver et défendre les conditions dans lesquelles ils ont obtenu l'accès aux données. Les limites et garanties relatives à ces données sont probablement plus restrictives que celles applicables à d'autres ensembles de données qu'ils ont le droit de consulter mais il y a une bonne raison à cela. Ils doivent donc les respecter.

C. Autres questions

80. On fait parfois valoir qu'il faudrait obtenir le consentement des répondants avant de diffuser les microdonnées à l'extérieur du service national de statistique. Ce genre de démarche devrait être découragé car:

- a) Elle soulève des problèmes pratiques importants;
- b) Les données fournies ne sont pas identifiables et ne servent qu'à un usage statistique, conformément à l'objectif de la collecte des données;

c) Il est très difficile de fournir la totalité des informations dont un répondant a besoin pour prendre une décision en toute connaissance de cause – et, par conséquent, bon nombre de répondants s’opposent à la diffusion de données les concernant, simplement à titre conservatoire. L’échantillon perdra rapidement de sa représentativité s’il repose uniquement sur les données fournies par les répondants qui ont accordé leur consentement.

Toutefois, il est obligatoire, comme indiqué ailleurs dans les présentes lignes directrices, de garantir la transparence des modalités mises en place. Ainsi, on peut faire valoir qu’un consentement passif doit être obtenu.

(Note: Un consentement donné en toute connaissance de cause conviendrait dans une situation où la publication de petits agrégats permettrait aux utilisateurs de déduire des conclusions concernant la situation d’une seule unité d’échantillonnage (personne ou entreprise, par exemple) prise en compte dans cet agrégat. La probabilité que le problème se pose est plus grande dans le cas des statistiques des entreprises.)

81. La question du consentement a aussi une autre facette. Les données d’un service national de statistique peuvent comprendre des informations recueillies directement par le service et des données collectées par les autorités administratives et communiquées au service national de statistique. Sauf disposition contraire de la législation ou d’un autre instrument pertinent, un service national de statistique ne devrait pas divulguer des données de source administrative sous la forme de microdonnées sans le consentement de l’autorité administrative intéressée (qui peut se trouver dans l’impossibilité de le donner en raison des engagements qu’elle a pris vis-à-vis des répondants). Même lorsque des données administratives se trouvent déjà dans le domaine public, la courtoisie veut que l’on avise les autorités administratives de leur diffusion éventuelle pour qu’elles aient l’occasion de formuler des observations.

82. Il importe que les services nationaux de statistique prévoient un dispositif spécifique dans le cas où l’accès aux microdonnées susciterait un débat public. Ils ne doivent pas exclure a priori l’éventualité qu’un tel débat se déclenche.

Quels sont les principaux arguments qu’un service national de statistique peut faire valoir?

a) Les services nationaux de statistique peuvent mettre l’accent sur le soin qu’ils apportent à la protection de la confidentialité de leurs données, notamment en anonymisant les microdonnées, en dressant une solide barrière de protection physique et en veillant à mettre en place un processus d’évaluation pour peser le pour et le contre entre les avantages conflictuels, pour le public, de la protection de la confidentialité, d’une part, et de la recherche, d’autre part;

b) S’il arrive qu’un chercheur commette une infraction et qu’un service national de statistique soit interrogé à ce propos, il devrait rendre publiques à la fois l’infraction commise et la sanction infligée; il devrait indiquer clairement que le responsable de l’infraction est le chercheur mais que le service national de statistique va prendre les mesures qui s’imposent pour sanctionner cette faute;

c) Les services nationaux de statistique devraient appeler l’attention sur l’intérêt public général qu’offre l’accès aux microdonnées, notamment dans le cas où une infraction a été commise;

d) Il faudrait solliciter le concours de personnes renommées et respectées qui sont prêtes à adhérer publiquement aux dispositions mises en place. L'aide de hauts fonctionnaires spécialisés en matière de confidentialité pourrait être particulièrement précieuse à cet égard.

VIII. PROBLÈMES DE GESTION ASSOCIÉS À LA PUBLICATION DES MICRODONNÉES

A. Gestion de la prise de décisions en matière de confidentialité

83. Il y a toujours un risque d'identification, même s'il est très faible. Il existe à présent des logiciels permettant d'évaluer la proportion d'enregistrements uniques en leur genre et présentant donc un risque d'identification.

84. C'est au chef du service de statistique ou à son représentant qu'il incombe de se prononcer au sujet de la publication d'un fichier de microdonnées, que ce soit par le biais d'un fichier de microdonnées anonymisé (à usage public ou sous licence), d'un moyen d'accès à distance, ou d'un laboratoire de données. Pour pouvoir prendre cette décision, le chef du service de statistique a besoin d'un avis lui permettant, par exemple, de savoir si:

a) Le risque d'identification est suffisamment faible;

b) Les ajustements apportés aux éléments de données n'ont pas exagérément altéré le fichier de microdonnées le rendant ainsi inutilisable aux fins de la recherche; et

c) Les variables qui ont été écrasées ont été choisies à bon escient, compte tenu à la fois des besoins des chercheurs et du risque d'identification.

85. Pour illustrer ce dernier point, des choix peuvent être faits entre le volume de détail fourni sur des variables conduisant à l'identification telles que les éléments d'information géographiques, l'âge et la structure du ménage.

86. Il faudrait mettre en place un dispositif pertinent pour donner un avis sur la question de manière cohérente. Un tel dispositif doit souvent être appuyé par des moyens de recherche appropriés et pourrait s'inscrire dans un cadre méthodologique. Les études de cas 16 et 17 décrivent respectivement les dispositifs mis en place en Slovénie et en Australie.

B. Gestion des métadonnées

87. Pour que les utilisateurs puissent exploiter efficacement les microdonnées, ils doivent avoir accès aux métadonnées pertinentes. Celles-ci comprendraient:

a) Une description de l'enquête et notamment toute information utile sur la qualité;

b) Une liste des éléments de données et des classifications utilisées (parfois dénommée «dictionnaire des données»); et

c) Des définitions des éléments de données.

La disposition visée sous a) garantira que les microdonnées ne seront pas utilisées si les données ne sont pas réellement bien adaptées à l'objectif recherché.

88. Comme les microdonnées sont fournies sur support électronique, les métadonnées devront être communiquées sous une forme accessible. Si possible, les métadonnées devraient être publiées en même temps que les microdonnées. La copie papier reste un moyen de diffusion efficace bien que le site Web du service national de statistique présente une utilité croissante à cette fin.

C. Gestion des infractions commises par le chercheur

89. Il faudrait s'attacher à réduire la probabilité des infractions telles que décrites dans le chapitre précédent. Néanmoins, des infractions peuvent survenir et il conviendrait de mettre en place des procédures pour y faire face.

90. Les infractions doivent être prises au sérieux. Sinon, le public perdra confiance dans le dispositif institué. En outre, les infractions se répéteront plus fréquemment si elles restent impunies.

91. Il y a plusieurs façons de réagir aux infractions. Par exemple, si une infraction à la loi a été commise, il faudrait envisager d'engager des poursuites, solution certes coûteuse mais essentielle pour démontrer l'importance que le service national de statistique attache à la confidentialité, et réduire la probabilité d'infractions futures.

92. En outre, il faudrait, au minimum, interdire aux chercheurs d'avoir accès ultérieurement aux microdonnées.

93. Il faudrait également envisager d'arrêter de communiquer des données à l'organisme dont relève le chercheur, tout au moins jusqu'à ce que:

a) Cet organisme ait pris les mesures qui s'imposent pour sanctionner l'infraction commise par le chercheur; et

b) Le service national de statistique ait la conviction que les mesures voulues ont été mises en place au sein de l'organisme pour réduire au minimum les risques d'infractions futures.

94. La communauté des chercheurs est généralement favorable à l'adoption de mesures strictes contre le nombre relativement restreint de chercheurs en infraction qui risquent d'entacher la réputation de la communauté des chercheurs. Il y va de leur intérêt à long terme.

95. Dans le cas des infractions mineures, un avertissement peut s'avérer suffisant.

IX. QUELQUES QUESTIONS PARTICULIÈRES

A. Accès international

96. Il importe de procéder à des comparaisons à l'échelon international pour bien comprendre l'efficacité des politiques et programmes adoptés par les différents pays. Les gouvernements, en particulier, jugent ces comparaisons utiles aux fins de l'évaluation des politiques. De toute

évidence, il y a avantage à permettre aux chercheurs qui participent à des comparaisons internationales, de même qu'aux organisations internationales, d'avoir accès à des microdonnées mais cela ne va pas sans risques (tels que la transmission des microdonnées sans autorisation). Sauf à EUROSTAT, les fonctionnaires des organisations internationales ne sont soumis à aucune législation nationale ou internationale, en dehors du règlement du personnel en vigueur dans l'organisation. Par conséquent, la prudence est de mise dans ce domaine. Le principal problème est lié à la plus grande difficulté qu'il y a à sanctionner les infractions commises par le personnel des organisations internationales ou des chercheurs vivant dans d'autres pays. Par ailleurs, la probabilité d'identification est beaucoup plus restreinte (tant que le chercheur ne transmet pas les microdonnées à une tierce personne dans le pays hôte). Un autre problème tient au fait que de nombreux pays ne sont pas légalement autorisés à communiquer des données aux organisations internationales ou aux chercheurs qui se trouvent en dehors de leur territoire.

97. Les présentes lignes directrices suggèrent que l'on remplace une stratégie d'évitement des risques par une démarche de gestion des risques dans le contexte de la communication de microdonnées. Les risques sont plus faibles ou perçus comme tels si l'organisme qui reçoit les données jouit d'une réputation de crédibilité et de fiabilité. Ils sont aussi plus faciles à justifier si le pays qui fournit les données tire avantage de l'étude pour laquelle les microdonnées sont communiquées notamment s'il s'agit d'une étude internationale entreprise par une organisation internationale ou d'un effort de recherche international réputé comme, par exemple, la LIS (Luxembourg Income Study).

98. La mondialisation accentue l'importance de ce genre d'études internationales. Les services nationaux de statistique devraient donc être en droit de soutenir les études en question en fournissant les microdonnées requises. Toutefois, les dispositions instaurées à cette fin devraient n'avoir qu'une fonction habilitante (en ce sens que le service national de statistique concerné devrait pouvoir décider dans chaque cas s'il convient ou non de fournir les données demandées) et prévoir aussi des garanties et des conditions de publication appropriées. Il faudrait que les services nationaux de statistique soient mieux préparés à communiquer des microdonnées dans les cas où les risques sont plus faibles et les avantages plus grands.

99. Quelles sont les diverses possibilités qui s'offrent aux chercheurs d'avoir accès aux ensembles de données d'autres pays? Comment les organisations internationales peuvent-elles obtenir l'accès aux microdonnées à des fins statistiques et de recherche? Les possibilités sont notamment les suivantes:

- a) Les données sont recueillies directement par l'organisation internationale (ou le chercheur) ou par le biais d'intermédiaires (un organisme d'enquête spécialisé, par exemple) de telle sorte que l'on fait savoir au moment de la collecte des données que les microdonnées vont être transmises aux fins de la recherche;
- b) Les fichiers à usage public lorsqu'il en existe;
- c) Les fichiers de microdonnées anonymisés sous licence, lorsque les pays sont en mesure d'en produire;
- d) Les moyens d'accès à distance assortis de mesures de protection appropriées; et

e) La collaboration avec un chercheur intégré au service national de statistique ou établi dans le pays où se situe le service national de statistique.

100. Du point de vue de l'accès aux microdonnées, les enquêtes de la catégorie a) sont préférables pour les chercheurs internationaux. L'étude PISA en constitue un bon exemple. Toutefois, de manière générale, ces collectes de données ne sont pas effectuées en vertu de la législation statistique en vigueur dans les divers pays. La qualité des microdonnées, et notamment les taux de réponse, pourraient en souffrir dans le cas de certaines études. Cela dépendra de la nature de l'étude et de la réputation de l'organisme qui l'entreprend, de même que de l'engagement contracté par les responsables de la collecte des données. Ce facteur doit être pris en considération par les chercheurs internationaux qu'ils optent ou non pour cette approche.

101. Il arrive que le service national de statistique puisse répondre aux besoins de l'étude en fournissant des données très détaillées à des fins d'analyse, mais pas de microdonnées. Cette démarche est suivie dans le programme de comparaisons internationales et l'étude OCDE/Eurostat sur les parités de pouvoir d'achat.

102. En ce qui concerne la démarche visée sous b), on ne dispose de fichiers à usage public que pour quelques pays. La constitution de fichiers de microdonnées anonymisés sous licence (en d'autres termes l'approche c)) peut être envisageable si elle n'est pas exclue par les règles du service national de statistique. Si les services nationaux de statistique sont autorisés à fournir des microdonnées par ce biais, les facteurs qui pourraient être pris en considération sont les suivants:

- a) La confiance qu'inspirent le chercheur et l'organisme dont il relève;
- b) Le degré d'importance de l'étude pour le pays; et
- c) La compatibilité ou non de la diffusion de ces données avec les engagements pris vis-à-vis des répondants au moment de la collecte des données.

103. Il serait probablement plus rassurant, pour bon nombre de pays, de communiquer des données à une organisation internationale spécifique ou pour un projet de recherche précis plutôt que de les diffuser de façon plus générale à la communauté internationale des chercheurs. En outre, il peut y avoir des conditions ne s'appliquant qu'à certains chercheurs en particulier. Par exemple, des pays peuvent ne se sentir en confiance pour communiquer des données aux chercheurs que s'ils passent par l'intermédiaire du service national de statistique dans le pays d'origine du chercheur. En tout état de cause, il serait de bonne règle de ne divulguer les données qu'à certaines conditions, soit en prenant un engagement, soit en signant un mémorandum d'accord. Parmi les conditions pourraient figurer:

- a) La limitation de l'accès aux données à certaines divisions des organisations internationales et l'interdiction de transmettre les données des tiers;
- b) La restriction des utilisations pour lesquelles les microdonnées peuvent être exploitées sans autorisation;
- c) L'obligation de rendre les microdonnées sur simple demande (par exemple si elles contiennent des erreurs);

d) La possibilité de faire des observations sur des publications s'appuyant sur les microdonnées; et

e) L'énonciation claire des conséquences qu'aurait le non-respect des conditions de diffusion.

104. Le moyen le plus efficace pour réagir au non-respect des conditions de diffusion consisterait à suspendre toute communication ultérieure de microdonnées. La question pourrait également être abordée avec de plus hauts fonctionnaires de l'organisme concerné. Dans le cas des organisations internationales, cela pourrait se faire par la voie diplomatique pour les infractions les plus graves. Toutefois, l'essentiel est de ne pas laisser des infractions se produire sans réagir. Sinon, elles se répéteront.

105. Pour de nombreux pays, l'utilisation de moyens d'accès à distance (c'est-à-dire la voie d'accès d)) pourrait être la meilleure solution pour permettre aux chercheurs internationaux d'obtenir des données. Dans ce genre de formule, un plus grand nombre de contrôles s'exerce et il est plus facile, en cas de contestation, de défendre la position des services nationaux de statistique concernant l'accès international aux microdonnées. Toutefois, l'applicabilité de ce genre de modalité à l'accès international doit encore être améliorée. L'expérimentation est importante.

106. Il existe encore une autre possibilité. Les chercheurs internationaux, notamment ceux qui œuvrent au sein des organisations internationales, pourraient travailler par le biais des réseaux de chercheurs nationaux pour atteindre leurs objectifs (autrement dit, selon l'approche e)). En effet, ces chercheurs nationaux pourraient mener leurs recherches dans les locaux du service national de statistique lorsqu'il s'agit d'études internationales particulièrement importantes.

107. Les services nationaux de statistique devront décider s'il convient ou non d'autoriser les chercheurs internationaux à avoir accès aux données, compte tenu de toute la gamme des questions examinées dans la présente section. Ils ne devraient pas perdre de vue qu'il faut encourager l'usage d'une méthode de gestion des risques. Dans certaines applications de la recherche, les avantages pourraient contrebalancer les risques, pour autant que l'arrangement mis en place soit légal. Les risques pourraient être plus limités pour certains organismes que d'autres. Les services nationaux de statistique devront également décider quel est le mode d'accès le plus approprié. Afin de favoriser la cohérence du processus décisionnel, les pays devraient élaborer, en ce qui concerne l'accès des chercheurs internationaux et organisations internationales aux microdonnées, des lignes directrices conformes à leur propre législation. Sinon, ils peuvent décider de modifier leur législation de manière à autoriser l'accès aux microdonnées dans les cas qui le justifient.

108. L'étude de cas 18 décrit les modalités employées dans l'étude PISA de l'OCDE.

B. Mise en corrélation des données

109. La mise en corrélation des ensembles de données, qu'elle se fasse par correspondance exacte ou par appariement statistique, peut créer une valeur ajoutée considérable. Elle permet d'élargir sensiblement la gamme des analyses. La recherche en matière de santé, en particulier, est un domaine où les ensembles de données corrélés peuvent revêtir une importance

particulière. Il est souhaitable que les services nationaux de statistique soient associés à la mise en corrélation des ensembles de données à des fins statistiques.

110. De plus en plus, les chercheurs s'attachent à utiliser des ensembles de données corrélés comprenant des liens avec les ensembles de données du service national de statistique ou d'autres organismes de statistique (y compris les données issues du recensement de la population dans certains pays). L'organisme de statistique doit assurer la garde de ces ensembles de données corrélés. Il arrive aussi qu'il soit jugé préférable de confier à l'organisme de statistique le soin de conserver les ensembles de données émanant d'autres entités, en raison des garanties de protection qu'il offre et de la confiance qu'il inspire au public.

111. Lorsqu'il y a manifestement avantage à établir une corrélation entre des données, une telle corrélation comporte aussi des risques notamment si l'entité qui a la garde du fichier corrélé n'a pas mis en place des mesures de protection de la confidentialité semblables à celles qui existent souvent au sein du service national de statistique. Les risques d'identification sont plus grands quand les ensembles de données sont corrélés. La manière de percevoir les choses est aussi importante. Les études réalisées dans bon nombre de pays mettent en évidence l'inquiétude que suscite dans le public la mise en corrélation de bases de données. Il importe particulièrement que les quatre principes énoncés dans le chapitre IV du présent document soient respectés dans le cas des ensembles de données corrélés.

112. Dans les pays où il existe des commissions chargées de la protection de la confidentialité ou des organes équivalents, ceux-ci devraient souscrire à l'adoption de modalités s'appliquant à la mise en corrélation des données.

113. Les études de cas 19 et 20 décrivent la manière dont le Canada et la Suède gèrent, respectivement, les modalités relatives à la mise en corrélation des données, à l'intérieur de leurs frontières.

X. LISTE DES ÉTUDES DE CAS

Annexe 1.1	Législation visant à appuyer la diffusion des microdonnées – Australie
Annexe 1.2	Législation visant à appuyer la diffusion des microdonnées – Finlande
Annexe 1.3	Cubes de données – Pays-Bas
Annexe 1.4	Usage public des microdonnées – Etats-Unis
Annexe 1.5	Diffusion de fichiers de microdonnées anonymisés (fichiers sous licence) – Australie
Annexe 1.6	Diffusion de fichiers de microdonnées sous licence – Pays-Bas
Annexe 1.7	Diffusion de fichiers de microdonnées sous licence – Suède
Annexe 1.8	Moyens d'accès à distance aux données – Canada
Annexe 1.9	Moyens d'accès à distance (pour l'accès aux microdonnées) – Australie

- Annexe 1.10 Accès à distance aux fichiers de microdonnées – Danemark
- Annexe 1.11 Programme du centre de données de recherche – Canada
- Annexe 1.12 Centres de données de recherche – Etats-Unis
- Annexe 1.13 Laboratoires de données – Pays-Bas
- Annexe 1.14 Accès aux microdonnées par le biais de laboratoires de données – Nouvelle-Zélande
- Annexe 1.15 Analyse des laboratoires de microdonnées – Italie
- Annexe 1.16 Gestion de la prise de décisions en matière de confidentialité – Slovénie
- Annexe 1.17 Gestion de la prise de décisions en matière de confidentialité – Australie
- Annexe 1.18 Étude PISA – OCDE
- Annexe 1.19 Gestion des projets de mise en corrélation des enregistrements de données – Canada
- Annexe 1.20 Mise en corrélation des données lors de la préparation des microdonnées destinées à la recherche – Suède
