



**Conseil économique
et social**

Distr.
GÉNÉRALE

CES/2003/7
7 avril 2003

FRANÇAIS
Original: ANGLAIS

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE POUR L'EUROPE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

Cinquante et unième réunion plénière
(Genève, 10-12 juin 2003)

**CONFIDENTIALITÉ DES DONNÉES STATISTIQUES
ET MICRODONNÉES**

Document de synthèse présenté par Statistics Sweden¹

I. INTRODUCTION

1. La principale difficulté que rencontre un institut national de la statistique concernant la confidentialité des statistiques et les microdonnées est de trouver un juste milieu entre protection du droit à cette confidentialité et utilisation accrue de microdonnées, qui permet de mieux répondre aux besoins des usagers. Cet équilibre est donc déterminant pour la statistique officielle qui devrait «... *constitu(er) un élément indispensable du système d'information d'une société démocratique, fournissant aux administrations publiques, au secteur économique et au public des données...*»². Par ailleurs, «*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*»³. Pour y parvenir, il est essentiel de combiner différentes mesures et actions et de traiter à la fois les aspects juridiques, techniques, administratifs et méthodologiques.

¹ Établi par Matti Niva, en collaboration avec Bo Sundgren et Ingrid Lyberg.

² Principes fondamentaux de la statistique officielle de l'ONU (art. 1), Commission de statistique de l'ONU, 1994.

³ Ibid., art. 6.

2. Le présent document, qui a pour objet de présenter certaines grandes questions liées à ces différents aspects, donne d'abord un aperçu des perspectives d'utilisation des microdonnées, traite ensuite de la confidentialité et des microdonnées puis de questions juridiques et enfin décrit brièvement les différentes démarches adoptées par des organismes concernant l'accès aux microdonnées. Cette structure correspond également plus ou moins à celle du séminaire.

II. UTILISATION DE MICRODONNÉES

3. Plusieurs faits nouveaux simultanés – avancées technologiques dans les matériels et les logiciels, documentation des données et utilisation de l'Internet – ont multiplié les possibilités d'utilisation des microdonnées pour la recherche. Les ordinateurs personnels modernes sont désormais suffisamment puissants pour traiter des séries volumineuses de données sophistiquées et les instituts nationaux de la statistique peuvent donc très facilement mettre ces séries à la disposition des chercheurs. Une part importante de leurs activités devrait d'ailleurs être de faire en sorte que les très nombreuses données stockées puissent être pleinement utilisées par les chercheurs et autres utilisateurs légitimes.

4. Jusqu'à présent, les statistiques agrégées étaient publiées en fonction de ce que l'institut national de la statistique jugeait important, même si les utilisateurs avaient bien sûr leur mot à dire. La présentation sous forme de tableaux imposait également des limites évidentes quant à la manière dont les statistiques officielles pouvaient être utilisées pour la recherche sociale et économique.

5. Ensuite pour accroître la valeur ajoutée de leurs données, les instituts nationaux de la statistique ont constitué des bases de données statistiques comportant des matrices de données agrégées, permettant ainsi aux usagers de compiler dans une large mesure leurs propres statistiques.

6. L'accès aux microdonnées représente un grand pas en avant puisque les chercheurs et les autres utilisateurs peuvent choisir les données dont ils ont besoin pour leur recherche. Cette innovation a eu également des incidences sur l'élaboration des théories dans le domaine de la recherche économique et sociale. De nombreux chercheurs peuvent se rendre compte de l'intérêt des microdonnées pour l'analyse des conséquences de certaines mesures de politique (voir, par exemple, Erikson, p. 2). L'explication théorique fondée sur des données agrégées peut ainsi être complétée par une analyse des mécanismes au niveau individuel.

7. Le fait de disposer de grands volumes de microdonnées longitudinales offre de nouvelles possibilités d'analyse. Par exemple, une correspondance entre différents fichiers de microdonnées portant sur plusieurs années permet de nouvelles analyses dynamiques. Ce type de recherche a été de plus en plus utilisé ces dernières décennies. Cette évolution est également évidente dans la recherche économique; dans le domaine de l'économie du marché du travail, par exemple, il est devenu habituel de comparer les données concernant les salariés et les entreprises pour analyser le marché du travail à la fois du point de vue de l'offre et de la demande (Westergaard-Nielsen, p. 2).

8. Les instituts nationaux de la statistique peuvent également intégrer plusieurs dossiers de microdonnées et créer de nouvelles bases de données. Toutefois, il faut habituellement un énorme travail statistique pour rendre la qualité des données acceptable. Statistics Sweden

a compilé certaines bases de données de ce genre. La base de données longitudinales «Louise» qui comporte des microdonnées anonymisées sur les individus et les familles, leur éducation, leurs revenus et leurs emplois pourrait servir d'exemple. Il faut ajouter qu'elle comporte depuis 1990 des données annuelles sur tous les adultes vivant en Suède et qu'elle est mise à jour chaque année. Une telle base de données intégrées offre de multiples possibilités d'analyses. On peut également relier plusieurs registres de microdonnées dans des bases ad hoc pour des besoins spécifiques.

9. La disponibilité accrue de microdonnées associée aux innovations dans le domaine des technologies de l'information a conduit à un nouveau type de démarche: l'extraction de données ou plus généralement la recherche du savoir dans des bases de données et l'extraction de données de bases multiples (Torra *et al.*).

10. Pour les instituts nationaux de la statistique, l'utilisation accrue de microdonnées est synonyme de valeur ajoutée en raison de la meilleure utilisation des données stockées. Elle devrait également améliorer leur légitimité vis-à-vis des répondants et du public en général et accroître la rentabilité des investissements réalisés dans les statistiques officielles.

III. CONFIDENTIALITÉ

11. L'une des grandes questions que pose l'utilisation de n'importe quelle microdonnée est la confidentialité. À vrai dire, toute utilisation de microdonnées, même anonymisées, comporte un risque. Même si la confidentialité n'a pratiquement jamais été violée dans les projets de recherche des instituts, sa protection reste et doit être prioritaire. Les résultats très positifs enregistrés jusqu'à maintenant sont en partie dus aux efforts déployés par les instituts nationaux de la statistique mais la raison pourrait bien être aussi que les chercheurs utilisant les microdonnées mettent en jeu leur propre capital humain. De plus, la question des microdonnées facilite généralement les contacts entre les instituts nationaux de la statistique et la communauté des chercheurs.

12. Le besoin de confidentialité et d'intégrité des données statistiques n'est pas une question nouvelle. Vincent P. Barabba avait déjà perçu le besoin de trouver ce juste milieu entre droit à la confidentialité et besoin accru d'information lorsqu'il déclarait en 1974 au nom du Bureau de recensement des États-Unis qu'il y avait un conflit inhérent à la collecte de données auprès de particuliers, un conflit entre le droit à la confidentialité de la personne, d'une part, et le recours par les autorités à des procédures contraignantes pour obtenir les informations dont elles ont besoin à des fins justifiées, d'autre part (Barabba, p. 34).

13. La communauté statistique internationale s'est penchée sur la question de la confidentialité et plusieurs documents ont été publiés. Dans une déclaration sur l'éthique professionnelle de 1985, l'Institut international de statistique (IIS) précisait que les données statistiques ne sont pas concernées par les identités individuelles, ce qui signifie que l'identité et les dossiers des sujets coopératifs (ou non coopératifs) doivent ... être confidentiels, que la confidentialité ait été ou non annoncée explicitement (IIS, 4.5 Maintaining confidentiality of records). En outre, les statisticiens devraient empêcher que leurs données soient publiées sous une forme qui permette de dévoiler ou de deviner l'identité d'un sujet quelconque (IIS, 4.6 Inhibiting disclosure of identities).

14. Les Principes fondamentaux de la statistique officielle de l'ONU sont également explicites sur ce point: «*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.*» (art. 6). Dans de nombreux pays, un code national de bonne pratique a été élaboré à l'intention de l'institut national de la statistique et habituellement la protection de la confidentialité en est l'un des éléments fondamentaux. C'est le cas du récent code de bonne pratique (National Statistics Code of Practice) de l'Institut national de la statistique du Royaume-Uni dans lequel le principe de la protection de la confidentialité constitue l'un des principaux engagements de l'Institut.

15. La protection de la confidentialité est de toute évidence un engagement primordial de tous les instituts nationaux de la statistique. Pour des raisons d'éthique, ils sont bien sûr concernés par l'intégrité des citoyens et par celle des entreprises mais leur engagement à protéger la confidentialité s'explique aussi par le fait qu'ils doivent être dignes de confiance s'ils veulent recueillir des données auprès des répondants.

16. Toutefois, compte tenu du fait que l'utilisation de microdonnées peut toujours mettre en danger la confidentialité (il n'existe pas de microdonnées complètement sûres), le véritable problème est de trouver un juste milieu entre informations plus nombreuses et confidentialité. Les instituts nationaux de la statistique estiment en général le risque lié à la divulgation de données dans différents domaines et les étapes du processus statistique pour différents types d'utilisation et tentent de minimiser ces risques. Ils ont d'ailleurs mis en place toute une gamme de méthodes.

17. Les caractéristiques spécifiques à la fourniture de microdonnées sur les entreprises par opposition aux données sur les particuliers soulèvent d'autres questions en matière de protection de la confidentialité. On se demande aussi si la protection de la confidentialité doit s'appliquer sans limites de temps (c'est-à-dire pour toujours) ou seulement pendant la durée d'une vie ou même moins?

18. Dans les pays d'Europe centrale et orientale, la situation institutionnelle et juridique en matière de statistiques a évolué considérablement au cours du processus de transition. Dans ces pays, la confidentialité des données peut présenter des caractéristiques spécifiques (en raison de la législation, de l'application du principe du flux unilatéral⁴, de la protection de la confidentialité de l'ensemble du système statistique). Une étude réalisée par le secrétariat de la CEE a conclu que des directives génériques concernant la confidentialité des statistiques seraient utiles (séminaire de la CES...).

⁴ Selon le principe du flux unilatéral, l'Institut national de la statistique a accès aux dossiers administratifs conservés par les ministères ou d'autres administrations publiques au niveau des microdonnées, alors que la possibilité d'un flux inverse de microdonnées protégées par la confidentialité statistique, qu'elles proviennent d'enquêtes statistiques ou d'une quelconque autre source, est totalement exclue.

19. Des événements récents ont rendu nécessaire un renforcement de la sécurité nationale, ce qui pourrait nuire à la confidentialité des statistiques. Si cela signifiait un accès illimité à des microdonnées réservées à des fins statistiques, alors la crédibilité de l'ensemble de la communauté statistique pourrait être remise en question.

20. Jusqu'à présent la question de la protection de la confidentialité se posait essentiellement en termes nationaux, mais dans le contexte européen elle devient du ressort des institutions de l'Union européenne comme Eurostat. De nouvelles possibilités mais aussi de nouvelles tensions apparaissent.

IV. QUESTIONS JURIDIQUES

21. Les mécanismes de protection de la confidentialité peuvent être fondés sur des textes juridiques et/ou des règles et réglementations appliqués par l'Institut national de la statistique. La situation en matière de législation varie d'un pays à l'autre et d'une région à l'autre. S'il existe une loi sur les statistiques relativement récente, elle contiendra habituellement des dispositions sur la communication des informations statistiques. En principe, les données recueillies à des fins statistiques, qu'elles soient collectées au titre d'une obligation imposée ou fournies à titre volontaire, ne peuvent également être utilisées que pour la production de statistiques. Toutefois, elles peuvent également être utilisées pour la recherche dans certaines conditions bien définies.

22. D'ordinaire, d'autres textes législatifs, tels que la loi sur les données à caractère personnel, s'appliquent à la production de statistiques et à la publication de microdonnées. Dans le cadre de l'Union européenne, la Directive du Parlement européen et du Conseil relative à la protection des personnes physiques (Directive 95/46 du 24 octobre 1995) est importante car elle renforce la protection juridique des personnes à l'égard du traitement des informations à caractère personnel (qu'elles soient automatisées ou contenues dans des fichiers structurés manuels) les concernant. Tous les États membres devraient avoir adopté une loi nationale sur les données à caractère personnel fondée sur la Directive de l'Union européenne, ce qui pourrait signifier qu'un organisme, comme par exemple une agence de contrôle des données, aurait le dernier mot concernant l'utilisation de ces données.

23. Parmi les textes juridiques portant uniquement sur la confidentialité des statistiques dans l'Union européenne, on trouve le Règlement du Conseil relatif à la statistique communautaire (n° 322/97) aux termes duquel tant les autorités nationales qu'Eurostat sont tenus de protéger les données confidentielles. Le principe de base est que les données confidentielles collectées exclusivement pour produire des statistiques communautaires doivent être utilisées seulement à des fins statistiques à moins que les répondants aient accepté de façon non ambiguë qu'elles soient utilisées à d'autres fins. Un règlement récent de la Commission, portant modalité d'application du Règlement du Conseil relatif à la statistique communautaire, concerne l'accès aux données confidentielles à des fins scientifiques (n° 83/2002) et porte sur les conditions régissant l'accès à des données confidentielles provenant de quatre enquêtes ou sources statistiques différentes de l'Union européenne. D'après ce texte, des chercheurs pourraient bénéficier d'un tel accès s'ils ont signé un contrat en réglementant les conditions avec Eurostat. Il reste à voir toutefois comment les États membres respectent dans la pratique cet arrangement juridique.

24. Concernant les données détenues par les autorités publiques, la situation en matière de protection de la confidentialité pourrait être bien différente. On peut défendre l'idée dans l'intérêt de la démocratie que toutes les informations qui sont créées à l'intérieur du secteur public doivent être publiques. Par conséquent, toutes les décisions prises par une autorité publique ainsi que les documents étayant ces décisions, y compris la correspondance échangée, devraient être rendus publics. Ce principe de la transparence de l'administration publique s'applique en Suède. Toutes les exceptions pour des raisons de confidentialité justifiées doivent être précisées dans des textes spécifiques comme une loi sur les statistiques.

V. DIFFÉRENTES MÉTHODES D'ACCÈS AUX MICRODONNÉES

25. L'une des principales difficultés que rencontrent les instituts nationaux de la statistique est d'offrir à l'utilisateur différents moyens d'accès aux microdonnées. En principe, cela est possible et c'est d'ordinaire l'institut national de la statistique lui-même qui doit trouver la solution pratique et adaptée compte tenu de l'environnement institutionnel et organisationnel. Toutefois, de nombreuses recherches comparatives de méthodologie ont été entreprises depuis des années.

26. L'une des premières méthodes d'accès contrôlé aux microdonnées consistait à compiler des fichiers de microdonnées anonymisées à l'intention du public. C'était la solution que Statistics Canada avait adoptée au début des années 70 (Boyko & Watkins, p. 3). Toutes les précautions sont prises avant de publier les fichiers pour réduire la probabilité de divulgation. Depuis le début de ce programme, plus de 350 fichiers ont été créés et plusieurs autres pays ont choisi des solutions analogues. Les fichiers ont été très utiles aux chercheurs des universités et aux administrations publiques. Parmi les problèmes que posait leur utilisation, on peut citer un coût relativement élevé, notamment pendant les années 80 et au début des années 90 (utilisation de gros ordinateurs, politique de fixation des prix) et le fait que le processus d'anonymisation diminuait la valeur des données (ibid., p. 5).

27. Dans de nombreux pays, la communication aux chercheurs et à d'autres utilisateurs autorisés, en dehors de l'Institut national de la statistique de microdonnées d'où toute référence à l'identité des répondants a été supprimée, constitue toujours la principale forme de diffusion. En 2002, il y a eu environ 200 cas chez Statistics Sweden et le nombre est en rapide augmentation. S'il s'agit de données détaillées provenant de fichiers, elles ne sont en fait pas anonymes. Il est évident qu'en pareil cas l'institut national doit se fonder pour approuver leur communication sur les textes de loi en vigueur ainsi que sur d'autres règlements précisant les règles de confidentialité. En Suède, toute tentative de retrouver l'identité des répondants est considérée comme un délit.

28. En raison de la sensibilité des microdonnées et de la possibilité de leur redonner une identité, dans de nombreux pays les instituts nationaux de la statistique ne permettent pas l'utilisation de microdonnées hors de leurs locaux et ont créé à cet effet un laboratoire pour les chercheurs. C'est le cas, par exemple, au Danemark (Access to Microdata in the Nordic Countries, p. 14). L'institut national de la statistique peut également mettre en place et administrer un centre extérieur, par exemple dans une université. Dans les deux cas, il lui est plus facile de vérifier qu'il n'y a pas atteinte au droit de confidentialité des données.

29. Une démarche encore plus prudente consiste à réserver l'utilisation des microdonnées exclusivement au personnel de l'Institut. Dans certains cas, le chercheur fera partie du personnel pendant une période déterminée, ce qui lui permet d'avoir accès aux microdonnées pour pouvoir mener à bien ses recherches. De même, l'Institut pourrait avoir comme politique de recruter des chercheurs pour pouvoir exploiter la multitude de microdonnées pour des clients externes.

30. Il devient plus courant toutefois de permettre à certains instituts de recherche d'avoir accès en ligne aux microdonnées anonymisées. C'est la solution qu'ont choisie l'Australie, le Danemark et le Portugal. Les séries de données disponibles en ligne peuvent être limitées en raison de leur caractère sensible et également modifiées selon les besoins et les orientations spécifiques des instituts de recherche concernés. Il est évident que les systèmes comme l'accès en ligne par l'Internet sont très appréciés des chercheurs. Ils pourraient également présenter de l'intérêt pour un nombre croissant d'instituts nationaux de la statistique étant donné qu'ils leur permettent de contrôler dans une certaine mesure l'utilisation des microdonnées.

31. La question des tarifs est également importante en matière d'accès à des microdonnées. Normalement, l'Institut national a déjà reçu un financement pour la majeure partie du travail de compilation et de mise à jour des fichiers. Il serait raisonnable qu'en pareil cas le prix à payer corresponde au coût supplémentaire lié à la communication des données, lequel pourra être bien sûr déterminé de différentes manières, selon les principes de calcul appliqué. Toutefois, la fixation des tarifs pourrait également s'inspirer d'autres principes, comme les prix pratiqués sur le marché ou même la gratuité dans certaines conditions.

VI. CONCLUSION

32. Le présent document a mis en évidence la nécessité de trouver un juste milieu entre protection de la confidentialité des données statistiques et amélioration de l'accès aux microdonnées. Il a également montré qu'il fallait associer un certain nombre de mesures juridiques, administratives et techniques pour atteindre cet objectif et qu'il y avait de nombreuses manières d'y parvenir. La communauté statistique pourrait également adopter une politique commune et convenir de principes fondamentaux concernant l'accès aux microdonnées. Il faut espérer que le séminaire de la Conférence contribuera à une meilleure compréhension des options qui s'offrent.

RÉFÉRENCES BIBLIOGRAPHIQUES

Access to Microdata in the Nordic Countries (2003) : Statistics Sweden.

Barabba Vincent P. (1974): *The Right of Privacy and the Need to Know*. Proceedings of the Social Statistics Section, American Statistical Association 33.

Boyko Ernie & Watkins Wendy (2002) *Safe Data, Safe Places: No Either/Or Solutions*. Communication présentée à l'occasion du 19^{ème} séminaire du CEIES intitulé «Innovative Solutions in Providing Access to Microdata».

CES Seminar on Data Confidentiality (2003). Note du secrétariat de la CEE.
CES/BUR.2003/27/Add.1.

Declaration on professional ethics (1985). Institut international de statistique.

Erikson Robert (2002): *The right to privacy and the right to information*. Communication présentée à l'occasion du 19^{ème} séminaire du CEIES intitulé «Innovative Solutions in Providing Access to Microdata».

National Statistics Code of Practice. Statement of Principles (2002). National Statistics, Royaume-Uni.

Perpétuo Fernanda (2002): *Statistical Information System for Researchers*. Communication présentée à l'occasion du 19^{ème} séminaire du CEIES intitulé «Innovative Solutions in Providing Access to Microdata».

Sundgren Bo (2001): *Statistical Microdata – Confidentiality Protection vs. Freedom of Information*. Réunion de travail conjointe CEE/Eurostat sur la confidentialité des données statistiques. Skopje, (ex-République yougoslave de Macédoine).

Torra Vincenç & Domingo-Ferrer Josep & Torres Àngel (2003): *Data Mining Methods for Linking Data from Several Sources*. Communication présentée à l'occasion de la réunion de travail conjointe CEE/Eurostat sur la confidentialité des données statistiques. Luxembourg, 7–9 avril 2003.

Principes fondamentaux de la statistique officielle de l'ONU (1994). Adoptés par la Commission de statistique de l'ONU.

Westergaard-Nielsen Niels (2002): *Linking employer-employee data*. Communication présentée à l'occasion du 19^{ème} séminaire du CEIES intitulé «Innovative Solutions in Providing Access to Microdata».
