



# 19<sup>e</sup> séminaire du CEIES

Solutions innovantes  
permettant l'accès  
aux microdonnées

Lisbonne, 26 et 27 septembre 2002



COMMISSION  
EUROPÉENNE



THÈME 1  
Statistiques  
générales

*Europe Direct est un service destiné à vous aider à trouver des réponses aux questions que vous vous posez sur l'Union européenne*

**Un nouveau numéro unique gratuit:**

**00 800 6 7 8 9 10 11**

De nombreuses autres informations sur l'Union européenne sont disponibles sur l'internet via le serveur Europa (<http://europa.eu.int>).

Luxembourg: Office des publications officielles des Communautés européennes, 2003

ISSN 1725-1354

ISBN 92-894-5294-3

© Communautés européennes, 2003

---

**1<sup>ère</sup> journée**  
**26 septembre 2002**

**9H00 INSCRIPTIONS**

**09H30 SÉANCE D'OUVERTURE**

**Accueil des participants**

*M. J. Lamel, vice-président du CEIES*

*M. S. Bacelar, INE, Portugal*

**Allocution d'ouverture:**

*M. P. Geary, président du sous-comité*

**Exposé introductif:**

*M. Ph. Nanopoulos, Commission européenne, Directeur, Eurostat*

**09H50(1) ÉTAT DE LA QUESTION: VUE D'ENSEMBLE DES POLITIQUES ET PRATIQUES EN MATIÈRE DE DIFFUSION DE MICRODONNÉES**

**Présidence: Mme F. Perpétuo, Instituto Nacional de Estatística, Portugal**

*M. G. Duncan, professeur de statistiques, Carnegie Mellon University, États-Unis*

**10H20 DISCUSSION OUVERTE**

**10H50 - 11H15 PAUSE CAFÉ/THÉ**

**11H15(2) QUESTIONS JURIDIQUES GÉNÉRALES**

*Mme B. Pettersson, Statistics Sweden*

*M. E. Chatzidoukakis, Commission européenne, Eurostat*

**11H45 DISCUSSION OUVERTE**

**12H30 - 14H00 PAUSE DÉJEUNER**

**14H00(3) BESOINS DE LA RECHERCHE ET DE LA POLITIQUE EN MATIÈRE DE MICRODONNÉES**

**Présidence: M. U. Trivellato, Università di Padova, Italy**

*M. R. Blundell, Études fiscales, économétrie, IFS*

*R. Erickson, Sociologie, liens avec le FSE*

*N. Vestergaard-Nielsen, Aarhus School of Business, Danemark*

*M. P. Sestito, Ministero del Lavoro e delle Politiche Sociali, Italy*

**15H30 - 16H00 PAUSE CAFÉ**

**16H00 DISCUSSION OUVERTE**

**17H00 FIN DES TRAVAUX DE LA 1<sup>ÈRE</sup> JOURNÉE**

**2<sup>ème</sup> journée**  
**27 septembre 2002**  
*(début des travaux: 9h00)*

**09H00(4) COMMENT DONNER ACCÈS AUX MICRODONNÉES**

*Présidence: M. L. Cox, National Center for Health statistics, U.S.A.*  
*Mme F. Perpétuo, Instituto Nacional de Estatística, Portugal*  
*M. J. Nobel, Centraal Bureau voor Statistiek, The Netherlands*  
*M. G. Lang, INSEE, France*

**10H00-10H30 PAUSE CAFÉ/THÉ**

*Mme W. Watkins, Data Centre, Carleton University Canada/M. E. Boyko, Statistics Canada*  
*M. K. Schurer, UK Data archive, University of Essex, Royaume-Uni*

**11H30-12H30 DISCUSSION OUVERTE**

**12H30-14H00 PAUSE DÉJEUNER**

**14H00(5) SOLUTIONS INNOVATIVES**

**Présidence: M. E. Boyko, Statistics Canada**  
*M. G. Wagner, université européenne Viadrina, Allemagne*  
*M. K. K. Skjåk, Norwegian Social Science Data Services, Norway*  
*M. O. Andersen, Statistics Denmark*

**15H00 DISCUSSION OUVERTE**

**15H30 RÉACTION D'EUROSTAT**

*M. Ph. Nanopoulos*

**15H45 RÉCAPITULATIF PAR LE PRÉSIDENT DU SOUS-COMITÉ**

*M. P. Geary*

**16H10 ALLOCUTION DE CLÔTURE PAR LE PRÉSIDENT**

*M. J. Lamel,*  
*vice-président du CEIES*

**16H15-16H45 PAUSE CAFÉ/THÉ**

**16H45 FIN DU SÉMINAIRE**

## Contenu

*Les actes des séminaires constituent un recueil de documents préparés par les orateurs avant le séminaire. Ils ne reprennent ni les débats ni les discussions des tables rondes ou des panels.*

*Les documents présentés et publiés ci-après ne représentent que les opinions de leurs auteurs et ne reflètent pas nécessairement une position officielle de leurs institutions ou organisations.*

### 1<sup>re</sup> journée

#### **Exposé introductif :**

*Document : M. Ph. Nanopoulos, Commission européenne, Directeur, Eurostat* ..... 09

#### **THÈME 1 — ÉTAT DE LA QUESTION: VUE D'ENSEMBLE DES POLITIQUES ET PRATIQUES EN MATIÈRE DE DIFFUSION DE MICRODONNÉES**

*Présidence: Mme F. Perpétuo, Instituto Nacional de Estatística, Portugal*

*Synthèse: M. G. Duncan, professeur de statistiques, Carnegie Mellon University, États-Unis.* 16

#### **THÈME 2 — QUESTIONS JURIDIQUES GÉNÉRALES**

*Présidence: Mme F. Perpétuo, Instituto Nacional de Estatística, Portugal*

*Synthèse: Mme B. Pettersson, Statistics Sweden* ..... 18

*Synthèse: M. E. Chatzidoukakis, Commission européenne, Eurostat* ..... 19

#### **THÈME 3 — BESOINS DE LA RECHERCHE ET DE LA POLITIQUE EN MATIÈRE DE MICRODONNÉES**

*Présidence: M. U. Trivellato, Università di Padova, Italy*

*Synthèse: M. R. Blundell, Études fiscales, économétrie, IFS* ..... 20

*Synthèse: M. R. Erickson, Sociologie, liens avec le FSE* ..... 22

*Synthèse: M. N. Vestergaard-Nielsen, Aarhus School of Business, Danemark* ..... 23

*Synthèse: M. P. Sestito, Ministero del Lavoro e delle Politiche Sociali, Italy* ..... 25

## 2e journée

### THÈME — COMMENT DONNER ACCÈS AUX MICRODONNÉES

<b>Présidence:</b> <i>M. L. Cox, National Center for Health statistics, U.S.A.</i>	
<i>Synthèse: Mme F. Perpétuo, Instituto Nacional de Estatística, Portugal</i> .....	27
<i>Synthèse: M. J. Nobel, Centraal Bureau voor Statistiek, The Netherlands</i> .....	28
<i>Synthèse: M. G. Lang, INSEE, France</i> .....	29
<i>Synthèse: Mme W. Watkins, Data Centre, Carleton University Canada/M. E. Boyko, Statistics Canada</i> .....	30
<i>Synthèse: M. K. Schurer, UK Data archive, University of Essex, Royaume-Uni</i> .....	31

### THÈME 5 — SOLUTIONS INNOVATIVES

<b>Présidence:</b> <i>M. E. Boyko, Statistics Canada</i>	
<i>Synthèse: M. G. Wagner, université européenne Viadrina, Allemagne</i> .....	32
<i>Synthèse: M. K. K. Skjåk, Norwegian Social Science Data Services, Norway</i> .....	34

### Réaction d'Eurostat

<i>Document : M. Ph. Nanopoulos</i> .....	36
---	----

### Récapitulatif par le président du sous-comité

<i>Document : M. P. Geary</i> .....	39
-------------------------------------	----

<b>Liste des participants</b> .....	41
-------------------------------------	----

---

# CONSÉQUENCES DU RÈGLEMENT N° 831/2002 RELATIF À LA STATISTIQUE COMMUNAUTAIRE EN CE QUI CONCERNE L'ACCÈS AUX DONNÉES CONFIDENTIELLES À DES FINS SCIENTIFIQUES

**Photius Nanopoulos**

Commission européenne, Directeur, Eurostat

## Séance d'ouverture

### Exposé introductif

#### 1. Introduction

Le nouveau règlement n° 831/2002 de la Commission concernant l'accès aux données confidentielles à des fins statistiques a été adopté le 17 mai et est entré en vigueur le 7 juin 2002. Cette adoption a constitué l'aboutissement d'un long processus de discussion, de négociation et de rédaction avec les 15 États membres (EM). Le Comité du secret statistique a examiné ce projet de règlement en février 2001. Le CEIES a discuté de ce sujet lors de la conférence commémorant son dixième anniversaire en avril 2001. Après d'autres négociations, une nouvelle version du règlement a été approuvée par le Comité du secret statistique en décembre 2001, dans le cadre d'une procédure de comitologie. Le chemin a été long, mais, aujourd'hui, les véritables travaux de mise en œuvre des dispositions du règlement ont débuté. Nous les mènerons en étroite collaboration avec les instituts nationaux de statistique, les États membres et la communauté de la recherche.

#### 2. Besoins de la recherche et types de données

La vie moderne se caractérise, entre autres, par l'intérêt grandissant et la demande croissante pour des stratégies basées sur des éléments concrets, des analyses politiques ainsi qu'un suivi des politiques et de leur impact. Ce type d'activités nécessite des informations actuelles et détaillées et requiert souvent des analyses plus approfondies que celles publiées aujourd'hui par les instituts de statistique. On considère parfois que ces analyses se situent en dehors des attributions des instituts nationaux de statistique (INS) ou même qu'elles constituent des activités susceptibles de compromettre l'indépendance perçue de ceux-ci. En effet, ces analyses sont souvent réalisées par des établissements universitaires ou des instituts de recherche indépendants.

Deux exemples de telles entités sont, au Royaume-Uni, l'institut (indépendant) pour les études fiscales (*Institute for Fiscal Studies*) et le nouveau *Centre for microdata methods and practice* (cemmap) – un projet commun de l'*Institute for Fiscal Studies* (IFS) et du département d'économie du *University College London* (qui fait partie de l'Université de Londres).

Des données détaillées sont nécessaires pour ce type d'analyses. Les données collectées et conservées par les INS sont considérées souvent comme la source la plus évidente et la plus intéressante. C'est la raison pour laquelle les INS et les autres organisations statistiques sont soumis à une pression de plus en plus forte leur demandant de fournir des données détaillées sur tout un ensemble de sujets. Pour l'Union européenne (UE) en particulier, les analyses et les recherches paneuropéennes prennent une importance croissante. Cette remarque vaut également pour la zone euro. Il est donc nécessaire d'avoir accès à des séries de données paneuropéennes pour mener ces travaux. Eurostat détient de nombreuses séries de données de ce type et c'est pourquoi il est considéré, par analogie à la situation nationale, comme la source potentielle naturelle, simple et directe pour celles-ci.

Dans le même temps, les organisations statistiques – les INS comme les institutions supranationales et internationales – considèrent de plus en plus qu'une utilisation plus intensive des données qu'elles détiennent

constitue un apport important à la société et une partie de leur obligation de mieux employer leurs ressources (données). Mais, il y a des limites à ce que les organisations statistiques, en particulier les INS, peuvent faire et à la façon dont elles peuvent intervenir. Le rôle des chercheurs et des établissements de recherche est donc essentiel et il s'accroît encore.

La nature de la recherche a également évolué. On ne peut satisfaire une recherche très moderne avec des données agrégées, il est nécessaire de disposer de microdonnées pour réaliser une analyse fine. Parallèlement à cela, il y a eu une évolution (“révolution” serait peut-être un terme plus approprié) de la capacité informatique de la recherche – dans le matériel comme dans les outils logiciels. Celle-ci a considérablement augmenté la demande d'accès aux enregistrements de microdonnées pour le calcul de matrices de corrélation, l'estimation de modèles et d'autres analyses qui dépendent du contexte du sujet de recherche.

### 3. Quels sont les problèmes?

Il existe toutefois, bien entendu, des problèmes dans la fourniture d'un accès à des données confidentielles en raison précisément de leur nature confidentielle. Diverses approches ont été suggérées et testées pour éviter ou réduire les problèmes. L'anonymisation des séries de microdonnées a été largement utilisée, notamment pour les données sur les ménages et les individus. Mais on sent bien que l'anonymisation ne suffit pas pour protéger totalement les données. Par “totalement”, on entend une protection à 100 % – tout pourcentage inférieur signifie qu'il y a un risque, si petit soit-il. Pour certains, toute valeur inférieure à 100 % est inacceptable. Il y a une corrélation entre le contenu d'informations de la série de données et le risque de divulgation. Le risque ne peut être réduit qu'en retirant des informations. Malheureusement, parfois, la meilleure information à retirer pour réduire le risque est précisément le détail auquel s'intéresse le chercheur. La recherche s'est penchée et se penche toujours sur l'évaluation des risques.

Quels sont les risques de réidentification par le chercheur? Tout d'abord, se pose la question de savoir si le chercheur peut voir réellement les enregistrements individuels. Certaines méthodes d'accès utilisées préviennent cela: une fois que les vérifications ont été effectuées, le chercheur ne reçoit que les résultats de l'analyse (un tableau, des coefficients ou un ensemble de paramètres). Si les protections sont bonnes, les risques sont faibles. De plus, les contrats régissant l'accès aux données passés avec le chercheur exigent habituellement de celui-ci qu'il n'essaie pas d'identifier les enregistrements ou affirme l'avoir fait. S'il le fait de manière accidentelle, il doit en informer uniquement l'organisation fournissant l'accès, de manière à ce que des améliorations puissent être apportées.

Il y a ensuite le problème de la réidentification par le biais des publications. Ici, les risques sont encore plus faibles, puisqu'il faut franchir trois obstacles: les protections de base mises en place pour l'accès et pour le contrôle des résultats de la recherche, la responsabilité du chercheur qui doit s'assurer que rien de confidentiel n'est publié et l'organisation qui vérifie les résultats avant leur publication. Pour certains, ces procédures peuvent sembler excessives, pour d'autres, elles constituent une exigence minimale. Mais pour toutes les fins pratiques, le risque peut être considéré comme très proche – même suffisamment proche – de zéro.

### 4. Quels types de données sont sensibles à la réidentification?

Pour peu que l'on dispose d'informations suffisamment détaillées, il est possible de réidentifier presque tout enregistrement. Le danger réside en partie dans les détails donnés dans l'enregistrement et en partie dans la rareté du sujet concerné dans la population étudiée. Bien entendu, les chercheurs souhaitent pour leurs travaux des données détaillées. Mais celles-ci peuvent être trop sensibles – d'où les mesures de protection destinées à réduire les détails accessibles au chercheur. La rareté du sujet – en particulier lorsque celui-ci est unique ou quasiment unique – est considérée comme un problème survenant principalement dans les enquêtes sur les entreprises. Celles-ci peuvent couvrir intégralement la population ou au moins la strate contenant ces éléments rares. Il n'y a donc même pas, dans ce cas-là, la protection dont on dispose habituellement dans les enquêtes sur les ménages ou les individus et qui résulte du fait que l'on ne sait pas si le sujet concerné a été inclus dans l'enquête. Les méthodes innovantes de génération de séries de microdonnées anonymisées pour les enquêtes sur les entreprises – des séries de données acceptables pour le chercheur en termes d'informations retenues et pour les INS et les sujets concernés en termes de préservation de la confidentialité – constituent donc un domaine de recherche passionnant et potentiellement très utile. Est-il possible de mettre ce système en place? Comment? C'est là le défi.

Le problème de la rareté des sujets apparaît également dans les enquêtes sur les individus et les ménages, mais on considère généralement que celle-ci est nettement moins problématique. La perte de détails résultant du regroupement de catégories ou des intervalles entre variables est souvent suffisante pour protéger l'identité des personnes concernées. Si nécessaire, on prend également d'autres mesures. Une méthode supplémentaire dans le cas d'enquêtes par sondage consiste à ne pas révéler que la personne a participé à l'enquête – il s'agit d'un moyen bien connu des INS et des autres instituts d'enquête.

### **5. De quelles méthodes disposons-nous pour le contrôle de la divulgation statistique (CDS)?**

Le développement de pratiques et de méthodes dans de nombreux INS a été largement pragmatique. Un ensemble de méthodes a été mis au point au fil des années, principalement pour les mises en tableaux – la production traditionnelle des offices statistiques. Parmi les méthodes utilisées à l'heure actuelle figurent:

- la réduction de la couverture géographique;
- l'arrondi;
- le regroupement ou la combinaison de catégories;
- l'ajout de bruit ou de perturbations;
- la permutation de données;
- la troncation par le haut (ou le bas) (top-coding/bottom-coding);
- l'imputation de valeurs à partir d'un modèle;
- la suppression de champs ou de cellules;
- l'obtention de la permission du répondant de divulguer les informations.

Il existe deux possibilités pour contrôler les résultats obtenus à partir d'analyses de données confidentielles: utiliser les mêmes méthodes pour vérifier les résultats tabulaires et créer des séries de microdonnées anonymisées. Il convient d'ajouter que le nouveau règlement 831/2002 nous donne à présent une définition juridique des séries de microdonnées anonymisées. Les "microdonnées anonymisées" sont "les fichiers statistiques individuels qui ont été modifiés afin de minimiser, selon les meilleures pratiques actuelles, le risque d'identification des unités statistiques auxquels ils se rapportent". Différentes techniques ont été proposées pour la création de séries de microdonnées anonymisées – il s'agit essentiellement des mêmes techniques que celles mentionnées plus haut, mais aussi de techniques de micro-agrégation et de suppression de variables. Une approche innovante suggérée récemment par Abowd et Woodcock est la création de séries de données synthétiques – conservant la structure interne des variables, mais ne contenant pas de données réelles.

De nombreux instituts statistiques considèrent la recherche dans ces domaines comme le moyen de parvenir à une meilleure compréhension de ce qui a été fait et de la manière de développer des méthodes plus efficaces. L'économie de ressources potentielle est en particulier importante si l'on automatise la vérification qui est effectuée actuellement – la mise en œuvre de nouvelles méthodes susceptibles d'être plus rapides et plus transparentes est toujours un aiguillon important pour les différentes organisations.

### **6. Conséquences du règlement n° 831/2002 pour Eurostat**

Tout d'abord, que fait le nouveau règlement 831/2002? Sur le principe, il établit les procédures et les conditions permettant d'accorder un accès à des données confidentielles à des fins scientifiques.

Le règlement se réfère à quatre sources importantes:

- le panel communautaire des ménages (PCM);
- l'enquête sur les forces de travail (EFT);
- l'enquête communautaire sur l'innovation (CIS);
- l'enquête sur la formation professionnelle continue (CVTS).

En résumé, les chercheurs doivent appartenir à des institutions de recherche et à des organismes établis dans les États membres. Une proposition détaillée doit être préparée, celle-ci doit indiquer l'objectif de la recherche et le détail des données à consulter. Les conditions et les protections sont fixées par contrat. Il n'y a aucun droit d'accès. Les EM peuvent retirer les données de leur pays pour tel ou tel projet de recherche. L'accès aux séries de données confidentielles peut intervenir dans les locaux d'Eurostat, avec des vérifications sur les données de sortie et les résultats afin de préserver la confidentialité, ou se faire par le biais de séries de microdonnées anonymisées.

Pour Eurostat, les conséquences du règlement et de sa mise en œuvre sont considérables. On dispose toutefois de précédents et d'expériences sur lesquels on peut s'appuyer. À Eurostat par exemple, on fait déjà appel à des dispositifs inspirés des "centres sécurisés" ou des "espaces sécurisés". Ceux-ci sont utilisés par les chercheurs travaillant sous contrat avec Eurostat. Le panel communautaire des ménages (PCM) a déjà ouvert la voie – à l'origine, en fournissant un accès contrôlé à des microdonnées confidentielles et, plus récemment, en créant et diffusant des séries de microdonnées anonymisées. Ces approches peuvent être développées et étendues aux autres enquêtes citées dans le règlement et à un plus grand nombre de chercheurs.

De nouvelles procédures pour recevoir les demandes de recherche, évaluer les chercheurs et leurs demandes ainsi qu'établir des contrats sont en cours d'élaboration. Cette interaction étroite avec la communauté scientifique, dans ces proportions et à cette fin, est relativement nouvelle pour Eurostat. En fin de compte, les mécanismes à mettre en œuvre au titre du règlement 831/2002 doivent être faciles à utiliser et fournir un service à la communauté de la recherche. Eurostat considère qu'il est très important de consulter celle-ci pour connaître ses exigences en termes de données et de mécanismes. Nous avons débuté avec le soutien du CEIES. Nous examinons le type de moyens dont disposent les chercheurs dans les États membres et ailleurs (la "Luxembourg Income Study", par exemple) ainsi que les attentes qui sont les leurs. Il nous faut également expliquer les contraintes et développer l'appréciation et l'acceptation de celles-ci.

## **7. Conséquences du règlement n° 831/2002 pour les États membres et leur relation avec Eurostat**

Le règlement invite Eurostat et les INS des États membres à travailler en étroite collaboration au développement d'un système permettant l'accès aux données confidentielles à des fins scientifiques. Cet ensemble d'activités très vaste – qui va de l'obtention d'accords sur la manière de vérifier et de protéger les résultats de la recherche, sur les protections et les contrôles des données et sur la façon de créer des séries de microdonnées anonymisées jusqu'aux procédures de traitement des demandes de recherche et de consultation. Néanmoins, la majeure partie de ces activités est probablement située en dehors du cadre thématique de notre séminaire. Les protections, contrôles et procédures feront appel à des approches et méthodes existantes – celles-ci sont les éléments dont nous disposons déjà. Ce séminaire s'intéressera aux approches innovantes qui pourront nous être utiles dans le futur.

L'un des aspects intéressants concerne le rôle des États membres par rapport aux demandes de recherche. Les EM doivent être informés de chaque projet de recherche. Toutefois, sous réserve que l'ensemble des conditions et exigences précisées ait été rempli par les chercheurs, on suppose, dans le règlement, que les EM accepteront de donner accès à leurs données.

## **8. Conséquences du règlement n° 831/2002 pour les chercheurs?**

Les conséquences du nouveau règlement pour la communauté de la recherche illustrent la nature du partenariat entre les offices statistiques et les chercheurs. D'une part, le règlement ouvre de nouvelles possibilités et, d'autre part, il impose une discipline et des limitations strictes, qui sont le prix à payer pour bénéficier de ces possibilités.

Tout d'abord, la communauté de la recherche doit accepter le fait qu'elle n'a pas de droit d'accès. Ensuite, les chercheurs devront accepter l'idée qu'il leur faudra respecter et faire respecter la confidentialité des données auxquelles ils ont accès. Les limitations et les protections peuvent être plus restrictives que celles qui sont en vigueur dans leurs universités ou auxquelles ils ont eu affaire avec d'autres ensembles de données, mais ils doivent s'y conformer. Les chercheurs devront accepter également qu'un autre organisme ait désormais le droit de poser des questions précises – non seulement sur les travaux et leurs objectifs, mais aussi, dans le cas de séries de microdonnées anonymisées, sur la façon dont les données seront conservées et dont l'accès à celles-ci sera contrôlé. De plus, après avoir accédé aux ensembles de données confidentielles, le chercheur devra présenter les résultats prévus afin qu'ils soient vérifiés avant leur publication ou leur parution sous une autre forme.

En retour, les chercheurs bénéficieront d'un accès plus simple aux séries de données couvrant l'ensemble des États membres. Jusqu'ici, il leur fallait, pour accéder aux données de chaque EM, passer par un long processus de dépôt de demande auprès de chaque État. Cette évolution permettra aux chercheurs d'effectuer des recherches et des analyses sur l'ensemble de l'Union européenne. Le règlement couvre quatre ensembles de données importants: on espère qu'à terme il sera également possible d'avoir accès à d'autres séries de données.

Ce règlement suscitera aussi une plus grande recherche en matière de méthodes de préservation de la confidentialité et d'évaluation des résultats des travaux menés à l'aide de données confidentielles. Un besoin de connaître le potentiel de divulgation des résultats issus de la modélisation commence, par exemple, à se manifester. Nous comprenons bien les problèmes de divulgation dans les données tabulaires – nous disposons pour cela d'une méthodologie, qui continue, en outre, à être développée –, mais nous avons une idée beaucoup moins précise des problèmes découlant de la modélisation et a fortiori de leurs solutions. Une restriction intuitive consiste à supprimer les informations sur les valeurs résiduelles – même si celles-ci sont d'un grand intérêt statistique pour les chercheurs – parce qu'elles donnent des informations sur les valeurs aberrantes qui sont souvent les sujets rares. Nous devons également en savoir plus sur le potentiel de divulgation des estimations de paramètres – notamment lorsque des modèles similaires sont exploités et comparés.

## 9. Conséquences du règlement n° 831/2002 pour les sujets concernés

Même si l'objet du règlement 831/2002 est d'améliorer l'accès aux données pour les chercheurs, il a également des conséquences pour les répondants qui ont fourni les informations initiales. Ces informations ont été données aux instituts statistiques de leur pays, dans le cadre d'une enquête volontaire ou obligatoire, ou bien elles ont été extraites de registres administratifs existants, dans le cadre d'une enquête statistique. Ensuite, l'institut statistique a transmis ces données à Eurostat après avoir supprimé les informations permettant une identification directe des sujets concernés. À cet égard apparaissent aussi d'autres conséquences pour les instituts statistiques. Le principe sous-tendant la collecte d'informations statistiques est celui du *consentement éclairé*. Selon celui-ci, le sujet concerné a le droit de savoir dans quel but ses informations seront utilisées et qui les consultera. L'argument ici est que, s'il existe une nouvelle dimension dans l'utilisation des données, le sujet doit en être tenu au courant.

Ainsi, le sujet concerné doit être informé, dans le cadre de l'enquête statistique, que les informations données seront utilisées à des fins statistiques et que celles-ci peuvent inclure des travaux de recherche effectués par des chercheurs externes, en plus de l'utilisation directe courante de l'institut statistique. Aux termes du règlement, les chercheurs peuvent appartenir à une institution d'un État membre et pas seulement à une institution de leur propre pays. À l'heure actuelle, les pratiques sont différentes d'un État à l'autre, de sorte qu'il est difficile d'indiquer ce qui devra être modifié. Dans la pratique, les choses qui devront changer seront peut-être peu nombreuses – les formulaires de consentement existants peuvent bien couvrir, implicitement, l'accès de chercheurs d'un autre pays à des fins de recherche statistique.

Il s'agit d'une question de nuance, évoluant le long de l'axe implicite-explicite avec l'aspect "éclairé" du consentement. Ce point peut nécessiter une recherche de terrain, y compris une étude qualitative auprès des sujets concernés. Il s'agit d'un élément important du contrat entre le sujet et l'institut statistique, que ce dernier considérera comme un facteur affectant le taux de réponse aux enquêtes volontaires.

## 10. Comment, selon nous, les activités évolueront-elles dans les deux à trois prochaines années?

Les instituts statistiques entretiennent, avec raison, une discipline stricte de préservation de la confidentialité des données qui leur sont confiées. Ils font preuve, à juste titre, d'une grande prudence en ce qui concerne les propositions visant à permettre un accès à ces données. Ils doivent être sûrs que l'action, quelle que soit, ne ruinerait pas ou n'anéantirait pas cette confiance ou la confidentialité des données. Une autre préoccupation se fait jour également: tout incident porterait préjudice à leur réputation (et à celle d'autres offices statistiques dans d'autres pays) et à la relation fondamentale mais sensible qu'ils entretiennent avec les fournisseurs d'informations, c'est-à-dire les sujets concernés. Le sentiment général dans la communauté statistique est qu'il faut avancer lentement, à petits pas. Par exemple, le *United States' Census Bureau* dispose aujourd'hui d'une chaîne de *Research Data Centres* situés dans ses locaux et dans plusieurs universités, où il est possible d'accéder à des microdonnées confidentielles. Mais il a commencé il y a plus de 20 ans et la route qui l'a amené jusqu'ici a été longue et est passée par le développement d'ensembles de données semblables aux séries de microdonnées anonymisées.

Dans les prochaines années, nous voyons les chercheurs accéder à des ensembles de données confidentielles via un "centre sécurisé" installé dans les locaux d'Eurostat et avoir, pour la recherche, un accès plus large à des séries de microdonnées anonymisées plus confidentielles. Ce dernier moyen d'accès est, à bien des égards, plus

facile et plus pratique – les chercheurs peuvent surtout travailler sur des microdonnées anonymisées dans le confort de leur propre bureau et institution. On sait toutefois que ces séries de données peuvent ne pas fournir l'ensemble des détails parfois nécessaires pour une recherche spécialisée. Et que certaines formes de données – notamment celles portant sur les entreprises – peuvent ne pas se prêter à la génération de microdonnées anonymisées utiles.

Le règlement 831/2002 envisage également, pour la recherche, un accès à des données confidentielles via des “centres sécurisés” situés dans les locaux des INS, dans les États membres. Ces centres auront les mêmes caractéristiques de sécurité et opéreront de la même manière que le centre implanté à Eurostat. La procédure de demande sera la même (quel que soit le centre utilisé), les restrictions imposées aux chercheurs et les protections de la confidentialité seront identiques.

Le règlement 831/2002 donne accès à quatre enquêtes importantes à des fins de recherche. Mais il ne s'agit là que d'un début. Lorsque l'on aura acquis une expérience avec celles-ci et que l'intérêt des chercheurs se sera accru, on pourra envisager de mettre à leur disposition davantage de séries de données. Il faudra probablement résoudre certaines questions juridiques, statistiques et méthodologiques, mais le souhait est d'utiliser autant que possible ces précieuses ressources en données.

Si l'on se projette beaucoup plus loin dans l'avenir, se pose la question de savoir si un accès recherche pour certaines formes de microdonnées peut être fourni à distance et de manière interactive via Internet ou la Grille (“the Grid”). La majeure partie de la réponse dépend de deux questions: la perception publique de la confidentialité peut-elle évoluer pour inclure ces aspects et pouvons-nous développer des méthodes permettant de contrôler et d'évaluer les risques impliqués? Des travaux seront nécessaires pour répondre à ces deux questions – en particulier à la dernière, qui pourra réclamer de nouveaux outils et de nouvelles approches.

Néanmoins, ce sont les États membres qui déterminent le rythme auquel nous pouvons avancer et ceux-ci sont, eux-mêmes, freinés par des questions juridiques et par la perception que le public et les répondants ont de la confidentialité, de la confiance et des risques.

## **11. Quels sont les domaines dans lesquels je souhaiterais que ce séminaire m'apporte des réponses?**

Je suivrai ce séminaire avec un intérêt particulier pour quatre domaines clés:

- les domaines de recherche spécifiques et détaillés que les chercheurs souhaitent étudier à l'aide de microdonnées – afin de déterminer les sources et les séries de données (ou les parties de celles-ci) dont l'accès deviendra une priorité;
- les rôles relatifs de l'accès via des “centres sécurisés” ou des “espaces sécurisés” et via des séries de microdonnées anonymisées, en gardant en tête les coûts relatifs et les situations dans lesquelles chacune de ces possibilités constitue l'option adaptée ou privilégiée;
- la réflexion actuelle sur les évolutions en matière de méthodes statistiques et de sécurité informatique qui peuvent permettre un accès à distance aux données; et, enfin et surtout,
- les méthodes innovantes, qui sont réalisables, ne représentent pas une charge excessive et peuvent être mises en œuvre rapidement pour nous permettre d'aider les chercheurs.

Évidemment, je pourrai ajouter des éléments à cette liste au cours de ces deux journées, après les présentations qui y seront faites.

## **12. Conclusions**

Nous – Eurostat, les INS, les chercheurs – avons besoin de méthodes complémentaires, nouvelles et innovantes. Ce séminaire est une excellente occasion d'écouter et d'apprendre et de repartir avec des idées et des suggestions. Ce séminaire est très important, car Eurostat et les INS doivent travailler avec les chercheurs. Le point essentiel, c'est la nécessité de faire le meilleur usage des données précieuses et détaillées dont dispose Eurostat. Cela requiert des méthodes éprouvées pour rendre acceptable l'accès aux données et pour quantifier les risques impliqués. Cela requiert également des idées, des outils, une recherche sur l'évaluation. Les chercheurs doivent montrer comment il est possible de donner accès aux données tout en préservant la confidentialité. La

---

relation entre l'accès aux données et la préservation de la confidentialité est souvent décrite comme un exercice d'équilibre. Mais il s'agit moins d'un acte d'équilibriste que d'une ouverture comportant une contrainte. Nous sommes probablement allés aussi loin que nous le pouvions avec les méthodes et les solutions acceptables aujourd'hui – la contrainte actuelle, c'est souvent le fait que les fournisseurs (les INS en premier lieu) ont des préoccupations et des contraintes d'ordre juridique, politique, psychologique et social, lorsqu'ils appliquent ces méthodes. Des solutions innovantes peuvent permettre de supprimer ou d'éviter ces contraintes pratiques et d'empêcher leur apparition. Et des méthodes innovantes peuvent apporter à Eurostat et aux INS des instruments supplémentaires qui leur permettront d'aider les chercheurs dans leur quête de données pour la recherche.

### **Référence**

Abowd J.-M. and Woodcock S.D. Disclosure limitation in longitudinal linked data. In Confidentiality, Disclosure, and Data Access, (pp. 215-278), North-Holland, 2001.

# POLITIQUE ET PRATIQUE EN MATIÈRE DE PUBLICATION DES MICRODONNÉES

**George T. Duncan**

Carnegie Mellon University, États-Unis

## Thème 1

### Résumé

Dans l'exercice de leurs responsabilités d'intendance, les offices statistiques doivent gérer la tension existant entre la garantie de la confidentialité et la fourniture d'un accès aux données (Duncan, Jabine & de Wolf 1993, Willenborg & de Waal 1996, Marsh *et al* 1991). Pour régler cette tension, il est nécessaire d'adopter des politiques permettant à un office statistique de diffuser des informations qui présentent:

- une *utilité des données*  $U$  élevée, une fidélité aux données originales aussi grande que possible dans les aspects fondamentaux (données analytiquement valides)
- un *risque de divulgation*  $R$  faible, de manière à ce que la confidentialité soit assurée (données sûres).

Ayant reconnu que la désidentification des données ne suffit généralement pas pour protéger leur confidentialité contre des attaques de "fouineurs" ("data snooper" – Elliot & Dale 1999), les instituts peuvent publier des microdonnées dans le cadre de politiques d'*accès restreint* ou diffuser des produits comportant des *données de nature confidentielle*.

Dans le cadre d'une politique et d'une pratique d'*accès restreint*, les procédures administratives imposent aux utilisateurs des conditions pour accéder aux données. Différentes politiques d'accès restreint (Jabine 1993a, b) ont été mises en œuvre. Aux termes de l'une d'entre elles, il peut être demandé à un utilisateur de données potentiel de solliciter le statut d'employé assermenté: il accepte par là d'être soumis à des conditions identiques à celles des employés permanents de l'institut. Si cette assimilation est approuvée, il peut lui être demandé de s'installer dans l'institut pour avoir accès aux données non confidentielles. Selon une autre politique, le chercheur doit déposer une caution. L'argent sera conservé si le chercheur ne respecte pas l'accord passé. Le partage non autorisé de données ou la réalisation d'analyses non précisées dans la proposition peuvent constituer des cas de non-respect de l'accord. Le *Census Research Data Center* de la *H. John Heinz III School of Public Policy and Management*, Carnegie Mellon University (voir <http://www.heinz.cmu.edu/census/>) est un exemple d'accord institutionnel donnant à des utilisateurs externes un accès restreint à des données. Grâce à cet accès à des données aussi intéressantes, le Centre a attiré des universitaires de renom national qui se sont engagés dans une recherche coopérative interdisciplinaire sur des questions politiques importantes.

Dans le cadre d'une politique et d'une pratique de diffusion de *données de nature confidentielle*, l'institut transforme les données originales pour diminuer le risque de divulgation. Pour les microdonnées, cette opération est effectuée par des techniques de limitation de la divulgation telles que (1) la publication d'un échantillon de données uniquement, (2) l'inclusion de données simulées, (3) le "brouillage" ("blurring") des données en les regroupant ou en ajoutant une erreur aléatoire, (4) l'exclusion de certains attributs, et (5) la permutation de données en échangeant uniquement les valeurs de quelques variables entre les sujets. L'abondance des méthodes existant pour limiter la divulgation statistique complique la tâche de l'institut, chacune ayant en effet des incidences différentes sur l'utilité des données et le risque de divulgation. Nous pouvons examiner de manière systématique l'impact simultané de la mise en œuvre de techniques de limitation de la divulgation et du choix de leurs paramètres sur ces deux valeurs. La mesure du risque de divulgation statistique  $R$  est une évaluation nu-

mérique du risque de divulgations non intentionnelles après la diffusion des données. La mesure de l'utilité des données  $U$  est une évaluation numérique de l'utilité des données publiées à des fins légitimes.

Les instituts ont besoin d'outils pour évaluer les politiques actuelles et futures en matière de publication des données de nature confidentielle. Le *diagramme de confidentialité R-U* ("*R-U confidentiality map*") est l'un de ces outils: il s'agit d'un graphique qui retrace l'incidence des modifications des paramètres d'une procédure de limitation de la divulgation sur le risque de divulgation  $R$  et l'utilité des données  $U$ . Les bases du diagramme de confidentialité R-U ont été présentées par Duncan et Fienberg (1999) et explorées plus en détail pour des données catégoriques par Duncan *et al.* (2001). Grâce à une représentation explicite de la façon dont les paramètres de la procédure de limitation de la divulgation influent sur  $R$  et  $U$ , l'interaction entre risque de divulgation et utilité des données apparaît clairement. Dans sa forme la plus basique, le diagramme de confidentialité R-U est un ensemble de valeurs couplées ( $R, U$ ) – risque de divulgation et utilité des données – qui correspondent à différentes stratégies de publication des données. Généralement, on applique, dans ces stratégies, une procédure de limitation de la divulgation, comme le masquage par l'ajout d'erreurs aléatoires. Ces procédures sont déterminées par des paramètres tels que, par exemple, l'ampleur de la variance d'erreur  $\lambda^2$  pour l'ajout de bruit. Une courbe est tracée dans le plan R-U, lorsque  $\lambda^2$  est modifiée. Visuellement, le diagramme de confidentialité R-U représente l'interaction entre le risque de divulgation et l'utilité des données à mesure que  $\lambda^2$  augmente et donc qu'un masquage plus important est imposé. Afin d'illustrer ce qu'elle peut apporter, la théorie du diagramme de confidentialité R-U est développée, dans le cadre de différents scénarios d'attaques de "fouineurs", pour un bruit additif appliqué à des données à une variable.

Un examen détaillé de la permutation de données figure dans le document. Dans cette méthode de limitation de la divulgation, certains champs d'un enregistrement sont remplacés par les champs correspondants d'un autre enregistrement. L'une des lacunes de la littérature actuelle réside dans le fait que celle-ci n'aborde pas les caractéristiques de performance de la permutation de données. Nous commençons à combler cette lacune en apportant des mesures du risque de divulgation ou de l'utilité des données et en examinant les interactions entre ces valeurs grâce au diagramme de confidentialité R-U.

Les instituts de statistique pourraient profiter des aspects suivants:

- Meilleure compréhension des risques empiriques de divulgation qui apparaissent lors de la diffusion de données.
- Quantification des pertes d'informations dues à l'application de procédures de limitation de la divulgation.
- Méthodes de diffusion des microdonnées avec des informations géographiques détaillées.

De nombreux domaines semblent prometteurs pour de futures recherches:

- Données virtuelles ou synthétiques;
- Accès en ligne aux bases de données statistiques;
- Méthodes permettant des données longitudinales;
- Utilisation des méthodes de Bayes.

## ACCÈS AUX MICRODONNÉES – LÉGISLATION EN VIGUEUR DANS LES PAYS SCANDINAVES

**Birgitta Pettersson**

Conseillère juridique, Statistics Sweden

### Thème 2

#### Résumé

Un groupe de travail scandinave a été créé pour examiner les questions de confidentialité liées à la fourniture d'accès aux microdonnées pour les chercheurs dans les pays nordiques. Dans le présent document sont abordés les grands principes de la législation en la matière dans ces pays. Il convient de noter que le groupe n'avait pas achevé son rapport au moment de la rédaction de ce document.

Dans l'ensemble des pays scandinaves, des lois sur les statistiques régissent l'utilisation des informations statistiques. De manière générale, les informations collectées à des fins statistiques ne peuvent être utilisées que pour la production de statistiques et la recherche. Néanmoins, dans le cas de Statistics Iceland, aucun texte spécifique ne traite de l'accès aux microdonnées. Les statistiques officielles islandaises s'appuient, au lieu de cela, sur la législation générale en matière de protection des données et, depuis plus récemment, sur la législation concernant la protection de la vie privée des individus.

Les législations nationales sur les données à caractère personnel s'appliquent également à la production de statistiques et à la publication de microdonnées. Ces législations contiennent des dispositions portant sur les exigences fondamentales en matière de traitement des données à caractère personnel. Ces critères incluent, entre autres, le fait que les données à caractère personnel ne peuvent être traitées qu'à des fins particulières explicitement définies et justifiées. Des règles très strictes s'appliquent pour le traitement de données à caractère personnel sensibles.

Des données – même anonymes – obtenues à des fins statistiques sont confidentielles. Des données confidentielles peuvent, en principe, être transmises à un tiers uniquement à des fins d'enquête statistique et de recherche scientifique. En vertu de ce grand principe, l'accès aux données peut être accordé sous des formes qui ne permettent pas l'identification directe ou indirecte des personnes ou encore l'identification indirecte d'autres entités concernées, comme les entreprises.

Si des microdonnées sont transmises à des chercheurs, les instituts nationaux de statistique (INS) peuvent imposer une restriction limitant le droit des chercheurs à utiliser les informations. Toutefois, en Suède, il n'est pas possible d'imposer de restrictions, lorsque les données sont communiquées à une autre autorité.

Même si la législation des pays nordiques est relativement semblable, les INS scandinaves ne diffusent pas les microdonnées de la même façon aux utilisateurs. En Norvège, en Finlande et en Suède, les séries de données sont fournies à des chercheurs qui travaillent en dehors des offices statistiques. Danmarks Statistik a une pratique différente. Étant donné que son principe fondamental est de ne pas laisser sortir les données, l'INS danois a créé un programme permettant à des chercheurs externes de travailler "sur site", dans ses locaux. Danmarks Statistik offre également à l'utilisateur un accès aux microdonnées à partir d'un ordinateur spécial situé dans ses services: l'utilisateur peut "piloter" cet ordinateur à partir de son propre bureau via une communication Internet cryptée.

---

## PROBLÈMES JURIDIQUES D'ORDRE GÉNÉRAL: LE POINT DE VUE D'EUROSTAT

**Efstratios Chatzidoukakis**

Commission européenne, chef de l'unité juridique d'Eurostat

### Thème 2

#### Résumé

1. Le principe de confidentialité a pour objectif de satisfaire la confiance des unités statistiques qui fournissent les données et dans le même temps, de garantir la fiabilité et la pertinence des informations statistiques. Les procédés technologiques modernes et les mutations socio-économiques mettent ce principe à l'épreuve. Parallèlement, il convient de répondre à la demande croissante des chercheurs qui souhaitent accéder aux données confidentielles à des fins scientifiques.
2. Eurostat se situe à un carrefour. Le service statistique de l'Union européenne doit respecter les exigences de confidentialité les plus rigoureuses des statistiques officielles et être fidèle à la confiance que lui accordent les autorités nationales, qui représentent ses principaux fournisseurs de données. L'évolution de la politique de l'Union européenne et les pressions exercées par la communauté scientifique afin d'avoir accès à des informations statistiques centralisées placent Eurostat dans une position unique.
3. Un cadre réglementaire bien établi permet à Eurostat de relever efficacement ces défis. Les règlements (CE) n° 1588/90 et n° 322/97 donnent notamment une définition objective de la confidentialité et fixent les conditions relatives à la transmission par les autorités nationales à Eurostat d'informations couvertes par le secret. L'accès aux données confidentielles, à des fins scientifiques, est reconnue comme un assouplissement légitime du principe en faveur de la communauté scientifique. Cet accès est autorisé dans des conditions clairement spécifiées, respectant les normes minimales du système statistique européen.
4. Le règlement (CE) n° 831/2002 récemment adopté par la Commission applique ces dispositions à quatre ensemble de données: le panel communautaire des ménages, l'enquête sur les forces de travail, l'enquête communautaire sur l'innovation et l'enquête sur la formation professionnelle continue. Les États membres et Eurostat travaillent en étroite collaboration afin d'assurer la mise en œuvre de ce règlement et de répondre aux attentes de la communauté scientifique. Le règlement représente un premier pas dans la bonne direction.

## LES BESOINS EN DONNÉES DANS UN MONDE VIEILLISSANT: MICRODONNÉES SUR LA SANTÉ, LE VIEILLISSEMENT ET LA RETRAITE

**Richard Blundell**

UCL et IFS

### Thème 3

#### Résumé

Ce document porte sur la collecte de données individuelles sur la santé, le vieillissement et la retraite. Il commence par établir le bien-fondé d'une conception pluridisciplinaire. Il examine les exigences précises concernant la santé, l'invalidité et la situation économique. Il aborde également le besoin de confidentialité et d'accès aux données administratives appariées, tout en plaidant en faveur d'un large accès à ces données pour la recherche. Des méthodes de référence pour garantir la confidentialité permettent aujourd'hui d'atteindre ces deux objectifs.

Dernièrement, les effets potentiels du vieillissement de la population ont été très largement reconnus dans les arènes politiques internationales. Cette prise de conscience s'est souvent exprimée par des préoccupations au sujet du paiement des retraites publiques, de la prestation des services sociaux et de santé publics et de l'impact économique d'une demande croissante de ce type de services, à mesure que la population vieillit. Il n'est plus pertinent d'assimiler la vieillesse soit à un âge biologique déterminé, soit à un changement social ou économique fixé par la loi, comme par exemple l'âge auquel un individu peut prétendre à une pension de retraite publique. Non seulement de nombreuses personnes choisissent de prendre leur retraite plus tôt, mais le recours de plus en plus fréquent à la préretraite plutôt qu'au licenciement a atténué également l'effet de ce qui était autrefois un phénomène de départs en retraite massifs à l'âge légal.

Cet agenda de la recherche, qui traverse le clivage entre la santé et les sciences sociales, se trouve confronté à quelques-unes des questions les plus centrales de la fourniture et de l'accès aux microdonnées. Les séries de données de ce type contiennent des informations sensibles sur la santé, les revenus, les biens et les interactions sociales. Elles comportent une grande variété de données quantitatives et qualitatives. Les liaisons avec les données administratives sont fondamentales. Les méthodes d'enquête requièrent la mesure précise de concepts très différents – qui vont des résultats en matière de santé jusqu'aux régimes de retraite. Cela représente un défi d'envergure pour les chercheurs, les experts chargés de l'élaboration des enquêtes et les pouvoirs publics. La santé, la retraite, les pensions et l'exclusion sociale jouent un rôle tel dans les débats politiques actuels en Europe, en Amérique du Nord et dans certaines parties de l'Asie que la disponibilité et la collecte de ces sources de données sont devenues des priorités de premier ordre.

La connaissance détaillée de l'évolution de la santé individuelle et de l'invalidité est un élément clé. La conséquence de l'accroissement de l'espérance de vie sur la morbidité dépend à la fois du taux d'évolution de l'incidence d'états invalidants et du taux de survie avec de telles invalidités. Alors que l'on a assisté à une controverse autour de la compression de la morbidité, les faits semblent indiquer que l'invalidité se réduit au fil du temps. Cet élément, combiné à l'observation d'une augmentation de la richesse de certains retraités, a conduit à reconnaître que certaines personnes âgées pourraient vivre une période d'épanouissement personnel plutôt que de déclin physique. Un problème scientifique et politique commun concernant les indicateurs du "vieillissement heureux" et la façon dont celui-ci peut être lié à la prévention des "évolutions défavorables" rend in-

dispensable de réaliser une enquête portant sur les changements clés et permettant d'étudier l'ensemble des dimensions du processus de vieillissement.

Le document cherche à démontrer que les données longitudinales sont essentielles pour comprendre la dynamique du vieillissement. Alors même que nous savons que le statut socio-économique avant la retraite influera à la fois sur la santé et la situation économique après la retraite, il nous est impossible de démêler les relations circulaires existant entre ceux-ci et d'autres facteurs sans disposer d'observations répétées sur les individus. Les comparaisons internationales sont également précieuses dans la mesure où elles accroissent la variation des résultats, des ressources individuelles et des systèmes institutionnels qui sont étudiés.

Une bonne base de données couvrira un large éventail de sujets, tous pertinents et très corrélés pour parvenir à une parfaite compréhension du processus de vieillissement – à la fois au niveau individuel et agrégé –, à l'adoption d'une approche pluridisciplinaire, à la représentation des perspectives sociales, psychologiques, biologiques et économiques. L'objectif partagé par tous doit être de répondre aux questions scientifiques et politiques, d'informer des évolutions dans ces deux domaines et de réunir ces deux sphères. Pour cela, il est essentiel de couvrir les questions clés, de réaliser des bases de données accessibles aux larges communautés de la politique et de la recherche et de produire des données qui puissent être utilisées pour des comparaisons internationales.

## LE DROIT AU RESPECT DE LA VIE PRIVÉE ET LE DROIT À L'INFORMATION

**Robert Erikson**

Swedish Council for Working Life and Social Research, Suède

### Thème 3

#### Résumé

Les citoyens vivant en démocratie ont droit à ce que leur sphère privée soit protégée, à ce que les informations les concernant demeurent confidentielles et à ce que leur situation personnelle ne soit pas exposée publiquement. Toutefois, pour juger l'action des pouvoirs publics et des autorités de l'État ainsi que pour pouvoir prendre des décisions politiques, ils ont également le droit de recevoir des informations sur les conditions, les tendances et les processus causatifs caractérisant leurs sociétés.

Il faut trouver un juste équilibre entre ces deux droits, ce qui suggère que les informations relatives à des individus identifiables devraient être protégées alors que les registres publics devraient être utilisés à des fins statistiques et scientifiques. Le fait que les données du domaine public puissent servir à des travaux de description et d'analyse permet également une utilisation efficace des investissements publics en matière de collecte des données.

Quatre exemples montrent comment des données stockées dans divers registres publics – recensement, archives hospitalières, registres des cancers et des décès, registres des abandons et des admissions scolaires et registres du chômage – ont pu être utilisées pour mettre en lumière les conditions et processus sociétaux. Dans toutes les études, les informations portant sur la situation individuelle des individus ont été tirées de registres, mais aucune donnée relative à des personnes identifiées n'a été dévoilée, que ce soit aux chercheurs ou à un public plus large.

Il semble donc tout à fait possible de protéger le droit d'un individu au respect de sa vie privée tout en garantissant son droit à recevoir des informations sur les conditions et les processus sociétaux. L'utilisation des données issues de registres publics à des fins statistiques et de recherche doit être limitée par certaines restrictions afin de protéger la vie privée des individus mentionnés dans les registres. Par conséquent, aucune information sur des personnes identifiées ne devrait être révélée et aucune action concernant les individus enregistrés ne devrait être entreprise en liaison avec les recherches effectuées. De cette façon, les deux principes - droit au respect de la vie privée et droit à l'information - sont respectés.

---

## L'APPARIEMENT DES DONNÉES EMPLOYEURS-SALARIÉS

**Niels Westergaard-Nielsen**

CLS, IZA et CCP, Aarhus School of Business, Danemark

### Thème 3

#### Résumé

##### *Introduction*

L'appariement des données employeurs-salariés permet de réaliser des analyses du côté demande du marché du travail. Depuis de nombreuses années, l'économie empirique du travail a pu analyser le côté offre de ce marché grâce aux répertoires et aux données issues d'enquêtes. Ces études ont produit d'énormes volumes de connaissances sur l'offre de travail, la structure salariale, l'impact de l'éducation, l'écart salarial entre les sexes, l'immigration et l'intégration, le chômage et l'emploi. Ces informations constituent aujourd'hui la majeure partie des connaissances dont nous disposons en économie et sociologie empiriques du travail, mais elles ne couvrent pas du tout les facteurs relatifs aux entreprises. Dans la plupart des analyses empiriques, l'entreprise est considérée comme une boîte opaque pour ce qui est des théories relatives à ce qui se passe en son sein. Dans les rares cas où l'on possède des éléments empiriques, il s'agit d'observations concernant une seule entreprise.

Les données appariées employeurs-salariés permettent d'étudier les différences entre les entreprises ainsi que les réactions à diverses interventions politiques. Elles permettent également de répondre à des questions d'intérêt stratégique sur les zones où sont créés et supprimés des emplois et sur la manière dont les différentes interventions politiques influent sur les entreprises et leur demande de main-d'œuvre. Ce phénomène a une dimension européenne, une dimension nationale et une dimension de gestion.

##### *Données*

L'étude des relations employeurs-salariés requiert des types de données spécifiques et des tailles d'échantillon différentes de celles des études traditionnelles situées côté offre. Le problème réside dans le fait que les individus "bougent" beaucoup entre les entreprises. Afin de pouvoir suivre les travailleurs d'un emploi à l'autre, les données doivent être relativement étendues. En outre, si l'on veut contrôler également les facteurs relatifs à la main-d'œuvre des entreprises vers lesquelles se dirigent les individus, on devra disposer de séries de données beaucoup plus importantes. Seul un ensemble complet d'observations permettra de suivre les individus d'un employeur à l'autre. Ce type de données est également le seul à autoriser une analyse complète de la mobilité entre les entreprises.

Les analyses employeurs-salariés peuvent en principe être basées sur les mêmes types de données que les informations sur les individus, c'est-à-dire sur des données d'enquête et des fichiers administratifs regroupés ou non en panels. On rencontre toutefois un certain nombre de problèmes spécifiques avec les enquêtes sur les entreprises. Les entreprises considèrent souvent qu'il faut consacrer trop de temps et d'efforts à celles-ci, ce qui peut entraîner une forte usure (les répondants ne répondent plus aux enquêtes). Il est évident que c'est dans le cadre d'un système de déclaration officiel, de procédures fiscales ou similaires que l'on recueille les données avec la meilleure qualité.

En outre, il existe un problème d'identité propre aux données employeurs-salariés, car les entreprises (à la différence des individus) fusionnent, se divisent, sont cédées ou restructurées.

Jusqu'à présent, peu de pays ont réussi à créer des données appariées employeurs-salariés. Les pays scandinaves bénéficient du fait qu'ils ont basé leur collecte de données sur des registres administratifs. La variable clé est, dans ce cas, un numéro personnel d'identification qui est utilisé dans l'ensemble du système administratif. Le numéro personnel d'identification assure le lien entre les registres. De même, les employeurs ont un numéro de registre. La fusion de ces informations permet de suivre les salariés par le biais des employeurs. En outre, le numéro de l'employeur peut être utilisé pour relier les informations concernant le personnel avec les statistiques des entreprises.

Au Danemark, ce type de données a été utilisé dans la recherche sur les créations et suppressions d'emplois et la rotation des travailleurs, dans les cas où les dimensions principales étaient la perspective macroéconomique et la relation entre la rotation du personnel et les cycles économiques. Nous sommes parvenus à un résultat commun dans l'ensemble de nos travaux: la classification des secteurs d'activité utilisée dans de nombreuses études est plutôt dépourvue de signification, car il existe des différences plus importantes entre les entreprises d'un même secteur qu'entre des secteurs d'activité différents.

Il faut respecter une condition pour utiliser ce type de données: les chercheurs doivent pouvoir travailler avec celles-ci de façon raisonnable.

#### *Accès aux données*

L'accès aux données a été largement examiné en relation avec les données individuelles. La confidentialité a toujours été une préoccupation majeure dans ce domaine de la recherche. Néanmoins, les fournisseurs de données ont, avec le temps, assoupli leurs règles en ce qui concerne la quantité d'informations à laquelle les chercheurs peuvent avoir accès à partir de leur propre ordinateur, pour réaliser leurs travaux. Premièrement, on a pris conscience qu'il était également de l'intérêt des chercheurs de ne pas divulguer des informations qui permettraient d'identifier des individus (tout leur capital humain est en jeu) et, deuxièmement, il existe de nombreuses séries de données bien connues que l'on peut télécharger ou acheter très bon marché. Troisièmement, l'expérience a montré que l'utilisation de données individuelles à des fins de recherche ne présente pas de risques importants. Quatrièmement, en ce qui concerne les données employeurs-salariés, il existe déjà des informations actualisées relativement détaillées sur les grandes entreprises dans le domaine public.

#### *Conclusion*

Les possibilités d'utiliser les données appariées employeurs-salariés sont grandes, car celles-ci peuvent être employées pour réaliser une recherche empirique sur ce qui se produit à l'intérieur d'une entreprise. Cela a des incidences sur notre compréhension des raisons qui font qu'une entreprise crée ou supprime des emplois et sur notre connaissance de la relation entre apport de main-d'œuvre et production du travail. Les résultats de la recherche comparative européenne pourraient avoir d'intéressantes perspectives et servir de fondement aux nouvelles politiques de l'emploi de l'UE. Elle présente également un potentiel important et jusqu'ici quasiment inutilisé pour la sélection des meilleures pratiques en matière de systèmes de gestion et de rémunération. L'Europe a donc, pour une fois, un avantage sur les États-Unis.

Dans le secteur médical, la plupart des citoyens et décideurs comprennent et soutiennent l'utilisation de données sensibles. Par le passé, il a été plus difficile d'obtenir un soutien en faveur de l'utilisation de données sur les individus et les entreprises, même si ce type de données est moins sensible. Évidemment, tous les types de recherche faisant appel à des données sensibles requièrent une politique de sécurité. Celle-ci devrait consister en une combinaison entre contrôle d'admission (sélection des personnes à qui l'accès peut être accordé) et limitations dans la matrice totale des informations avec lesquelles les chercheurs peuvent travailler au même moment. Néanmoins, il faudrait prendre conscience du fait que tous les mécanismes nécessitent une certaine dose de confiance dans le chercheur, puisque qu'il est impossible de développer un système de sécurité qui puisse prévenir totalement les divulgations intentionnelles. Des officiers de police ou des agents administratifs ayant accès à des enregistrements individuels ont déjà été jugés coupables de divulgations, mais cela n'a jamais été le cas de chercheurs.

Selon moi, le véritable danger avec les données basées sur les registres réside dans le fait qu'elles ne sont pas utilisées pour améliorer les conditions de vie de la population.

## BESOINS EN DONNÉES POUR L'ÉVALUATION DES POLITIQUES: QUELQUES RÉFLEXIONS

**Paolo Sestito<sup>1</sup>**

Banca d'Italia, Rome, Italie

Ministère du travail et des politiques sociales, Rome, Italie

### Thème 3

#### Résumé

L'évaluation des politiques est un domaine de plus en plus important pour l'analyse économique et statistique. Cette évaluation est, au fond, une analyse hypothético-déductive: son objectif est de comparer le résultat et/ou le comportement (de groupes) d'individus avec et sans l'application d'une politique donnée. Cependant, dans la pratique, il est impossible d'observer un même groupe de personnes dans deux environnements différant l'un de l'autre uniquement par la mise en œuvre ou non d'une politique. Les évaluations empiriques sont donc effectuées à l'aide d'une combinaison d'instruments économétriques et expérimentaux visant à comparer des groupes d'individus "semblables" l'un à l'autre.

Le document cherche à démontrer que trop d'énergies sont consacrées à l'élaboration de techniques économétriques et statistiques – en recourant à des hypothèses souvent arbitraires et non explicitées sur la distribution de phénomènes non observables – alors que trop peu d'énergies sont employées à améliorer la qualité, l'accès effectif et l'utilisation des données auxquelles ces techniques doivent être appliquées. Cette affirmation, qui n'est pas nouvelle, sera présentée en adoptant la perspective d'un décideur politique.

Une telle perspective est particulière. Le décideur s'intéresse surtout à l'examen de l'*impact différentiel des politiques* sur les individus et les contextes économiques locaux et à l'*efficacité de certains aspects de la mise en œuvre des politiques*. Dans le même temps, il doit examiner l'*impact de la politique sur l'équilibre général*, en tenant compte également des incidences qu'elle a pour les groupes qui ne sont pas directement concernés par l'intervention politique.

Ces particularités mettent en évidence l'importance de disposer de données plus nombreuses et de meilleure qualité, c'est-à-dire de données plus précises – en ce qui concerne notamment la mise en œuvre de la politique –, d'une plus grande désagrégation entre les différents groupes socio-démographiques et contextes locaux et d'une dimension temporelle suffisante pour saisir les effets sur l'équilibre général. Des priorités spécifiques sont dérivées de ces particularités. On préconise une *élaboration des données administratives déterminée par l'évaluation des politiques* et un *suivi de la mise en œuvre des politiques basé sur des sources multiples*, incluant l'utilisation d'indices de satisfaction des consommateurs issus de sources statistiques plus traditionnelles. L'*intégration de sources de données différentes* constitue une troisième priorité, mise en avant en particulier pour l'exploitation des sources de données administratives. L'*identification et le suivi continu des contextes économiques locaux* forment un quatrième élément. Enfin, le document met un *accent particulier sur les informations longitudinales*, qui permettent d'identifier les effets à long terme (et sur l'équilibre général) de la politique.

Pour finir, le document aborde brièvement les accords institutionnels envisageables susceptibles à la fois de favoriser une telle évolution en ce qui concerne le contenu et la qualité des données et des systèmes d'informa-

<sup>1</sup> Les avis exprimés ici sont purement personnels. Ils ne reflètent pas nécessairement ceux des institutions auxquelles j'appartiens.

tion et de faciliter l'accès aux données. Ce dernier point est, affirme-t-on, essentiel pour améliorer la responsabilité politique. La nécessité de renforcer la "voix" des organes chargés du suivi et de l'évaluation dans la conception des bases de données administratives et des systèmes d'information est un premier aspect mis en avant. La capacité d'influer sur l'élaboration concrète de ces systèmes serait grandement améliorée si l'on intégrait cette activité de conception dans la mise en œuvre concrète de la politique. Cependant, rassembler les activités de suivi et d'évaluation et la responsabilité de mise en œuvre d'une politique créerait un conflit d'intérêts insoluble. Le principe de la séparation entre "contrôleur" et "contrôlé" est donc un préalable indispensable à un suivi fructueux. Pour ce qui est de l'évaluation, la mise en pratique de ce principe de séparation par le recours à des appels d'offres ouverts pour les études d'évaluation présente ses propres avantages, mais fait apparaître aussi quelques problèmes. L'adjudication d'études d'évaluation individuelles garantit l'indépendance de l'évaluateur par rapport à l'entité juridique responsable de la mise en œuvre de la politique examinée. Néanmoins, ce système d'appel d'offres intervient très souvent à un stade relativement tardif et l'évaluateur désigné n'a pas la possibilité de contribuer à l'élaboration ex-ante d'un système d'information adapté. De plus, avoir pour évaluateur une entité désignée officiellement, même si elle a été sélectionnée selon des procédures transparentes de passation de marché public, ne garantit pas la conduite d'une réflexion appropriée sur la politique, une telle réflexion nécessitant en effet une véritable opposition de vues. Dans la solution préconisée, le recours aux appels d'offres pour les études d'évaluation officielles est complété par la production de fichiers à usage (quasi-)public, mis, sous certaines conditions, à disposition des chercheurs externes qui souhaitent réaliser leurs propres études d'évaluation.

Dans un système de ce type, l'entité en charge du suivi et de l'évaluation doit disposer d'un degré d'indépendance suffisant par rapport à celle qui gère les différentes politiques. L'organe qui met sur pied le suivi devrait avoir (au moins) quatre tâches spécifiques: a) contribuer à l'élaboration du système d'information, en intégrant les séries de données administratives entre elles et avec d'autres sources; b) faire rapport sur les politiques mises en œuvre suivant des normes statistiques appliquées de manière systématique et universelle; c) réaliser, en interne et/ou en recourant à des évaluateurs indépendants sélectionnés par des appels d'offres ouverts, des études d'évaluation officielles spécifiques, d) accorder aux tiers souhaitant réaliser leur propre analyse un accès aux bases de données utilisées pour l'évaluation officielle.

---

## SYSTÈME D'INFORMATION STATISTIQUE POUR LES CHERCHEURS

**Fernanda Perpétuo**

Instituto Nacional de Estatística, Portugal

### Thème 4

#### Résumé

À la fin de 2000 et durant 2001, l'INE du Portugal (INE-PT) a créé, sur le réseau privé du ministère des sciences et de la technologie (MCT), un site sécurisé qui fait actuellement l'objet d'essais, dans le but de mettre des données statistiques à la disposition des chercheurs accrédités par l'Observatoire des sciences et de la technologie (OCT). Ce site fournira des microdonnées anonymisées et des macrodonnées au niveau de la strate ainsi que les métadonnées respectives jugées pertinentes pour la recherche, l'interprétation, l'analyse et la réutilisation des données disponibles. Le présent document expose les exigences du protocole et les objectifs du système créé, explique comment le contenu est géré, énumère les principaux problèmes et indique les résultats projetés de l'évolution.

## LE CHANT DES SIRÈNES

**Joris Nobel**

Centraal Bureau voor Statistiek, Pays-Bas

### Thème 4

#### Résumé

Les statisticiens officiels entendent le chant séduisant des sirènes: des chercheurs de toutes sortes, dans les universités comme dans les administrations, veulent avoir accès aux microdonnées officielles. Toutefois, dans l'intérêt des répondants et des autres fournisseurs de données, un tel accès ne peut être accordé sans condition. Nous avançons entre Charybde et Scylla (Dublin 1992) dans notre Odyssée des données (Amsterdam 2001). Dans ma présentation, j'exposerai l'itinéraire choisi par le vaisseau néerlandais.

Si les années 80 peuvent être qualifiées de période de guerre froide entre l'Office statistique des Pays-Bas (CBS - Centraal Bureau voor Statistiek) et les milieux universitaires, les années 90 passent, quant à elles, pour une période de réconciliation et de pacification.

En 1994, un contrat a été élaboré et signé par CBS et l'Organisation néerlandaise pour la recherche scientifique (NWO): celui-ci prévoit la diffusion de fichiers standard de microdonnées tirées de l'enquête sociale.

En 1996, la loi relative à l'Office central et à la Commission centrale de la statistique confère à CBS l'autorité juridique de publier, sous certaines conditions, des fichiers standard de microdonnées tirées de l'enquête sociale.

En 1998, CBS a ouvert Cerem, son Centre de recherche sur les microdonnées, dans le cadre d'une expérience sur trois ans.

Aujourd'hui, l'accès aux microdonnées à des fins de recherche figure de nouveau à l'ordre du jour aux Pays-Bas. Des circonstances nouvelles réclament des politiques davantage tournées vers l'action.

Au sein de CBS, de nouveaux modèles de collecte et de traitement des microdonnées, qui se concentrent de plus en plus sur la mise en relation des fichiers administratifs, tendent à rendre obsolète le concept de fichier standard de microdonnées.

Les administrations néerlandaises expriment un besoin croissant d'accéder, à des fins de recherche, à des fichiers (administratifs) bien fournis, afin de leur permettre d'élaborer des stratégies et d'exercer leur responsabilité politique.

Les milieux universitaires néerlandais ont atteint une certaine saturation avec les fichiers standard: ils souhaitent désormais des fichiers plus riches. Les chercheurs de pointe se tournent soit vers une collecte "à la carte" de données primaires aux Pays-Bas, soit vers une analyse secondaire de fichiers de microdonnées américains.

Dans les années à venir, des modifications interviendront dans la législation statistique nationale et de nouveaux moyens d'accès aux microdonnées verront le jour. Ces deux types de modifications auront pour but d'améliorer l'accessibilité en toute sécurité aux fichiers de microdonnées à des fins de recherche. La législation devrait permettre la mise en œuvre d'une politique plus rationnelle pour l'Office statistique néerlandais. Parmi ces nouveaux moyens figurent l'extension des dispositifs sur site, mais surtout le développement des opérations à distance. La coopération internationale sur ces questions est une nécessité pour CBS, qui se trouve dans une société relativement petite et ouverte. La coopération avec les milieux universitaires néerlandais est prometteuse, mais fragile.

---

# LE SECRET STATISTIQUE ET LE COMITE FRANÇAIS DU SECRET STATISTIQUE CONCERNANT LES ENTREPRISES

## Gérard Lang

Le Chef de la Division, Environnement Juridique de la Statistique, INSEE

### Thème 4

#### Résumé

Le droit français fait une nette distinction, pour l'application du secret statistique aux données individuelles, entre le cas des données relatives aux personnes physiques et le cas des données relatives aux entreprises et aux autres personnes morales.

Si le secret des données statistiques individuelles relatives aux personnes physiques est absolu, celui qui s'applique aux données statistiques individuelles relatives aux entreprises l'est beaucoup moins. La loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques indique :

“Les renseignements individuels d'ordre économique ou financier figurant sur les questionnaires [statistiques publiques] ne peuvent en aucun cas être utilisés à des fins de contrôle fiscal ou de répression économique”.

En dehors des finalités interdites par la loi, l'accès aux données statistiques individuelles des entreprises n'est donc pas illégal.

C'est pour mettre en place la jurisprudence et les règles d'application indispensables que le décret n° 84-628 du 17 juillet 1984 d'application de la loi du 7 juin 1951 a mis en place, par ses articles 21 et 22, le comité du secret statistique concernant les entreprises. “Celui-ci est chargé d'étudier les problèmes que pose l'application des règles relatives [au secret statistiques concernant les entreprises]”. Après avis du comité, les ministres concernés peuvent prendre des décisions autorisant, par dérogation au secret des données individuelles, l'accès à ces données pour certaines finalités précises, dont la recherche scientifique. Ces données rendues accessibles comprennent presque toujours les données d'identification des entreprises et des établissements.

Le comité, qui se réunit trois fois par an depuis 1985 a examiné 1185 demandes d'accès à des enquêtes statistiques entre 1985 et 2001. La procédure, qui comporte une phase écrite et une phase orale, examine chaque demande au cas par cas.

Le comité, présidé par un haut magistrat, comprend des représentants des ministères (dont celui de la justice), des représentants du patronat et des syndicats de salariés ainsi que des représentant des utilisateurs.

On obtient ainsi une participation de la société civile aux décisions de dérogation au secret statistique qui permet un “contrôle démocratique” dont on ne trouve pas d'équivalent dans les procédures mises en place par le règlement n° 831/2002 de la Commission du 17 mai 2002 qui donne une base juridique permettant l'accès des chercheurs européens aux données confidentielles transmises à Eurostat par les Etats membres de l'Union européenne.

## **SÉCURITÉ DES DONNÉES, SÉCURITÉ DES ENVIRONNEMENTS: LES DEUX SONT INDISPENSABLES**

**Ernie Boyko, Statistique Canada**

Wendy Watkins, Carleton University

### **Thème 4**

#### **Résumé**

Le document présente une approche non technique pour résoudre le problème de la fourniture d'un accès aux microdonnées pour la recherche scientifique. Au lieu d'examiner les procédures utilisées dans ces situations, nous avons choisi d'évaluer les différentes méthodes de fourniture d'accès aux microdonnées employées au Canada. Nous traitons également les questions de confidentialité et joignons une note d'accompagnement qui aborde les points statistiques.

Le Canada est dans la position unique de pouvoir évaluer l'impact d'une modification de sa politique de diffusion des données en ce qui concerne les fichiers de microdonnées à grande diffusion (FMGD). Avec la mise en œuvre de l'Initiative de démocratisation des données, nous sommes passés d'une diffusion sporadique de ces fichiers à une situation où la quasi-totalité des universitaires du pays a accès à la plupart des archives de FMGD de Statistique Canada.

Nous avons tenté de décrire chaque méthode et d'exposer ses forces et ses faiblesses, telles qu'elles nous apparaissent. Nous examinons la situation des FMGD, des Centres de données de recherche, de la soumission de tâches à distance (accès à distance), du programme des chercheurs en résidence et des totalisations personnalisées.

Notre conclusion est donc que les FMGD offrent le meilleur rapport qualité/prix. Dans les cas où ceux-ci ne peuvent être utilisés, nous classons les autres méthodes selon leur contribution potentielle à la recherche scientifique.

---

## DES SOLUTIONS POUR PERMETTRE L'ACCÈS AUX MICRODONNÉES: L'EXPÉRIENCE DU UK DATA ARCHIVE

### **Professeur K. Schürer**

Directeur, UK Data Archive,  
Université d'Essex, Colchester, Royaume-Uni.

### **Thème 4**

#### **Résumé**

Le présent document examine la fourniture d'accès aux microdonnées sur la base de l'expérience du *UK Data Archive* (UKDA – Service d'archivage des données du Royaume-Uni), établi à l'Université d'Essex. Le UKDA est l'une des plus anciennes et des plus importantes archives de données de sciences sociales en Europe. Depuis plus de vingt-cinq ans, il transmet des microdonnées issues d'enquêtes produites par les administrations publiques aux universitaires britanniques sans avoir rencontré de problèmes de divulgation statistique en ce qui concerne la confidentialité des informations fournies par les répondants.

Le document aborde plusieurs thèmes. Tout d'abord, il traite de plusieurs questions générales concernant le partage et la diffusion de données à l'ère électronique. Cela semble indiquer que l'augmentation des sites Web de fourniture de données a fait naître l'idée que les microdonnées sont aujourd'hui peut-être plus exposées au risque de divulgation qu'elles ne l'étaient auparavant. Le document expose ensuite les exigences des chercheurs (principalement des universitaires) en matière de données. Il examine, dans un troisième temps, les avantages généraux du partage et de l'archivage des informations pour les producteurs de données. Cette partie comprend également une analyse de la nécessité de veiller à ce que les procédures adéquates de préservation des données soient en place, une situation que l'on maîtrise au mieux en recourant à la gestion des fichiers et à l'archivage des données. Les préoccupations des producteurs de données sont présentées dans une quatrième partie. L'expérience du UKDA en matière de diffusion des données comme dans son rôle d'intermédiaire entre les producteurs et les utilisateurs de données est ensuite décrite en détail. Cette vue d'ensemble fournit des informations sur l'ampleur des activités menées par le UKDA, sur les procédures qu'il a instauré pour vérifier et valider les données et sur la nécessité de créer une relation de travail réciproque entre l'archiviste et le producteur de données. Les méthodes de contrôle d'accès aux données et de garantie de la confidentialité établies par le UKDA sont décrites dans un sixième temps. Celles-ci consistent en une validation des données, en un suivi et un contrôle des utilisateurs ainsi que, essentiellement, en un accès contrôlé par le biais d'un système d'enregistrement des utilisateurs, une spécification des conditions d'utilisation des données et la mise en œuvre d'une unité de contrôle d'accès. La dernière partie du document est consacrée aux évolutions techniques en matière d'application de méthodes de divulgation statistique. Elle présente en particulier le développement d'un module de contrôle interactif de la divulgation statistique, dont le but est d'assurer la fourniture en ligne sans risque de microdonnées et de données tabulaires agrégées via une unité de contrôle d'accès sur le Web, qui permet aux utilisateurs de rechercher, de tabuler, de visualiser et de télécharger des données en ligne. On peut donc considérer que les technologies Web, plutôt que de constituer une menace pour les microdonnées, peuvent être exploitées pour créer un environnement sûr, surveillé et contrôlé pour la diffusion de celles-ci.

# DES SOLUTIONS INNOVANTES POUR PERMETTRE UN ACCÈS AUX MICRODONNÉES: L'EXEMPLE DU PANEL SOCIO-ÉCONOMIQUE ALLEMAND

**Gert G. Wagner**

Technische Universität Berlin (TUB)

et

DIW Berlin (Institut allemand de recherche économique), Allemagne

## Thème 5

### Résumé

Parce que l'élément central du paradigme scientifique est la reproduction et la réanalyse de résultats, il n'est pas suffisant de donner à un chercheur particulier un accès exclusif à une base de microdonnées. Tous les autres scientifiques qui sont intéressés par la même question de recherche (c'est-à-dire ses concurrents) doivent avoir, eux aussi, un accès équivalent à ces microdonnées. Du point de vue de l'opinion publique, la réanalyse est nécessaire pour éviter des erreurs qui pourraient avoir des conséquences indésirables si les décisions politiques s'appuyaient sur des résultats empiriques erronés. Ces éléments expliquent la nécessité d'accorder un accès aux microdonnées statistiques aux producteurs de données officiels mais aussi scientifiques. D'un point de vue juridique, il faut tenir compte non seulement des droits des répondants, qui réclament le respect de la confidentialité, et des scientifiques, qui souhaitent faire de la recherche, mais aussi d'un troisième droit, l'intérêt du grand public, qui demande des analyses statistiques valides et fiables qui serviront de bases aux conclusions stratégiques.

Pour évaluer le risque de divulgation d'un répondant, il est intéressant de souligner que l'on ne connaît pas un seul cas d'utilisation abusive de microdonnées *statistiques* par des institutions non gouvernementales, y compris des entreprises privées. Il est facile d'en comprendre la raison: ni les chercheurs universitaires, ni les organismes de recherche privés n'ont intérêt à ce que les données des répondants soient divulguées. La valeur réelle d'enregistrements isolés de données d'enquête, surtout d'échantillons représentatifs d'un pays tout entier, est presque nulle.

Néanmoins, tant qu'un élément d'information est présent dans un ensemble de données, il existe un risque de désanonymisation du répondant. En ce qui concerne la confidentialité, il n'y aura jamais de "séries de données totalement sécurisées" (pour lesquelles le risque de divulgation sera réellement égal à zéro). Les "espaces sécurisés" utilisés pour les analyses de données sont donc juste aussi importants que les techniques de non-divulgation des séries de données. Il existe différentes manières d'établir des espaces sécurisés: par des restrictions "technologiques", par des mesures juridiques – qui sont beaucoup plus "puissantes" – (comme la Loi fédérale relative à la protection des données [*Bundesdatenschutzgesetz – BDSG*]) et, demain, par l'adoption d'un code déontologique obligatoire.

Le document présente ces différentes solutions en s'appuyant sur l'accès aux microdonnées du panel socio-économique allemand (Sozio-ökonomisches Panel – SOEP).

La première solution consiste à créer un "espace sécurisé" en recourant à des "restrictions technologiques". Analyser les données dans le cadre d'un "Centre de données de recherche" (Research Data Center – RDC) est une possibilité directe offerte par de telles restrictions. On met en place un RDC de ce type pour analyser la correspondance "ultra sensible" des données du SOEP avec les données du géocode.

Un “cadre sécurisé” peut être créé également en appliquant des sanctions juridiques sévères en cas de violation des règles de protection des données. Les sanctions sont une mesure commune pour inciter à l’acceptation des lois. Il est difficile de comprendre pourquoi la protection des données devrait faire exception lorsque l’on réclame une sécurité totale. Pour les utilisateurs du SOEP, les sanctions en cas de violation des règles de protection des données sont fixées par le biais d’un contrat privé, étant donné que la loi relative à la protection des données (*BDSG*) n’est pas directement applicable, car les données anonymes ne sont pas de véritables données à caractère personnel protégées par la loi.

En généralisant l’exemple du SOEP, la communauté internationale des statisticiens pourrait renforcer sa position d’un point de vue juridique en établissant un “code déontologique” obligatoire pour les statisticiens et les organismes qui travaillent avec des microdonnées. Un tel code faciliterait grandement le “quotidien” des statisticiens.

La protection des données fait partie intégrante de l’éthique professionnelle des médecins (ainsi que de celle d’autres professions comme les avocats, les conseillers fiscaux, etc.). Il est nécessaire d’avoir, outre “l’éthique professionnelle”, une législation garantissant le “droit de maintenir secrètes des informations privées”, même si l’État, en la personne d’un procureur par exemple, demande ces informations.

Un tel code déontologique obligatoire doit faire partie de l’enseignement dispensé aux statisticiens et donc des examens qu’ils doivent passer. Un diplôme portant, entre autres, sur la “statistique” pourrait être ainsi un “permis d’analyse des microdonnées”.

## RÉDUIRE LES BARRIÈRES ENTRE LES UTILISATEURS ET LES DONNÉES DIFFUSION DES DONNÉES DE L'ENQUÊTE SOCIALE EUROPÉENNE

**Knut K. Skjåk**  
et  
**Kirstine Kolsrud**

Norwegian Social Science Data Services, Norvège

### Thème 5

#### Résumé

Le document traite essentiellement de la façon dont le Service norvégien de données sur les sciences sociales (NSD), qui servira d'archive à l'enquête sociale européenne (ESE), tentera de réduire les obstacles existant entre les utilisateurs et les données à l'aide de la technologie moderne.

L'ESE est une enquête initiée par les milieux universitaires dans le but principal de mesurer et d'expliquer la continuité et l'évolution dans trois grands domaines: les valeurs des individus, les orientations culturelles et nationales et la structure sociale sous-jacente de la société. La première enquête sera conduite dans plus de 20 pays européens entre septembre et décembre 2002.

L'ESE est le résultat d'un effort collectif réalisé par des chercheurs européens, des fondations scientifiques et des organismes de financement nationaux. Les données seront considérées et traitées comme un bien collectif. Les coûts et les efforts engagés collectivement tiendront un rôle essentiel dans la production des données. Aucune discrimination ne sera faite parmi les utilisateurs, ni par la tarification, ni par une politique d'accès sélective.

La nature même de l'ESE nous a déjà permis d'éliminer les deux premiers des trois principaux groupes d'obstacles – "juridiques et institutionnels", "culturels" et "liés à l'organisation des données" – rencontrés par des utilisateurs analytiques de microdonnées. On peut donc s'orienter, pour la diffusion des données, vers le troisième obstacle lié à l'"organisation des données". En substance, surmonter cet obstacle signifie produire des données faciles à exploiter pour l'utilisateur et les diffuser dans des systèmes conviviaux facilement accessibles pour les utilisateurs.

Afin d'aider les équipes nationales à produire des données comparatives normalisées de haute qualité, le NSD a créé un site Web intra-projet, un système qui fournit à l'ensemble des participants des documents, normes, définitions de données et autres outils de traitement de données. Nous pensons que ce système contribuera grandement aux efforts d'élimination des obstacles liés à la qualité des données pour les utilisateurs de l'ESE.

Le NSD a décidé de diffuser les données de l'ESE par le biais de deux systèmes basés sur le Web. Le premier, NESSTAR, est un système qui permet aux utilisateurs de:

- localiser des sources de données multiples dans les archives de plusieurs entrepôts de données;
- rechercher des métadonnées détaillées (documentation sur les données);
- rechercher et visualiser les données en ligne;
- télécharger les sous-ensembles de données pertinents dans l'un des nombreux formats disponibles.

Le système NESSTAR intègre des données et des métadonnées normalisées (DDI). Une fois qu'un ensemble de données est localisé et ouvert pour la recherche, les métadonnées peuvent être visualisées dans un navigateur combiné métadonnées/données. Les métadonnées sont affichées sous forme d'une arborescence extensible et l'utilisateur peut alors consulter des informations sur la série de données ou sur des variables particulières.

Dans cet outil, l'utilisateur peut passer facilement de la description en texte intégral des variables à une analyse statistique de base. Parmi les fonctions offertes par le navigateur en ligne, figurent, outre la possibilité de décomposer les données par variable et valeur et d'ajouter une pondération à n'importe quelle analyse, l'affichage tabulaire et graphique des données, la génération de tableaux à entrées multiples, les ventilations, la corrélation et la régression.

Les données de l'ESE seront également accessibles à partir d'un site Web dédié et ouvert, hébergé sur le serveur du NSD. En plus d'être un portail vers les données, le site Web des données de l'ESE (ESS Data Web Site) sera l'entrepôt des données et des documents de l'ESE 2002. Ce site accueillera tous les documents électroniques produits ou déposés dans les archives, y compris les données brutes, les programmes de données, les résumés techniques nationaux, les données finales et leur documentation. Avec des navigateurs Web traditionnels, les utilisateurs pourront consulter et télécharger:

- les séries de données en différents formats;
- les normes pour l'encodage des variables enregistrées oralement;
- le protocole de données ESE (définitions des variables, procédures de nettoyage, etc.);
- le rapport technique et le code;
- les autres documents, par exemple les questionnaires et rapports méthodologiques sur les données;
- une bibliographie ESE.

Une autre application de l'ESE sera de fournir aux étudiants en recherche comparative des ressources de formation basées sur Internet reposant sur les données collectées à chaque cycle du projet. Le nouveau programme de formation, qui sera accessible via des navigateurs traditionnels connectés à Internet, sera développé et géré par le NSD en étroite collaboration avec les experts des domaines couverts par les ressources.

## 19<sup>e</sup> SÉMINAIRE CEIES RÉPONSE D'EUROSTAT

L'objectif du séminaire, qui était l'échange de vues entre statisticiens et chercheurs dans le but de trouver un accès pratique aux informations confidentielles, a été, selon moi, parfaitement atteint.

Les discussions des deux journées ont fait apparaître une série de questions qui devront être examinées prochainement.

### 1. Les utilisateurs, c'est-à-dire la communauté de la recherche

- a) Les chercheurs veulent des données réelles et non des données perturbées ou corrompues.
- b) Les chercheurs veulent un accès facile aux données. Ils préféreraient avoir les données "à domicile" ou disposer d'un accès à distance via Internet.
- c) Les chercheurs veulent pouvoir combiner des informations de différentes sources, lier des relevés issus de différents registres.
- d) Les chercheurs veulent que l'accès soit gratuit ou soumis à des droits réduits.

Dans le même temps, les chercheurs sont prêts

- à assumer des responsabilités juridiques et personnelles pour assurer la confidentialité des données; et
- à accepter d'autres conditions, comme l'obligation de fournir des résultats publics et l'acceptation de certains contrôles.

Les premières questions sont les suivantes: "qui sont les chercheurs?", "quel est l'objectif de leurs travaux?", "quel est leur 'code de conduite'?". Pour nous, statisticiens, les réponses ne sont pas claires.

### *Le rôle de la communauté de la recherche*

La recherche socio-économique est plus importante que jamais pour le progrès et la survie de nos sociétés. Les quelques exemples présentés au cours de cette conférence ont illustré non seulement la nécessité de disposer de meilleures données, mais aussi les problèmes qui sont abordés. La plupart d'entre eux sont liés directement aux décisions politiques capitales qui devront être prises dans les années à venir.

Il est donc nécessaire de procéder à une planification afin de prévoir les besoins et de collaborer avec l'ensemble des parties concernées, de manière à ce que chacun soit prêt.

Des structures organisationnelles sont requises pour mener un bon dialogue et, au besoin, exercer une pression. Il convient d'étudier les possibilités de jouer un rôle qu'aura le FSE.

Il faut comprendre que l'ouverture de l'accès a un coût qui doit être pris en charge par ceux qui ont commandé la recherche. Nous devons sans conteste faire la distinction entre les différents objets de la recherche.

### 2. Les statisticiens officiels

Les statisticiens officiels sont confrontés à toute une série de problèmes.

Il est clair que nous sommes loin de pouvoir satisfaire les quatre demandes reprises plus haut et ce, pour quatre raisons:

- a) Un environnement juridique complexe, restrictif et peu clair. La mise en œuvre de dispositions juridiques trop restrictives et incohérentes donne aux utilisateurs l'impression d'un cadre "paranoïaque". Nous devons travailler en collaboration et réviser l'infrastructure juridique.

- b) Nos ressources ne nous permettent pas de progresser et de nous organiser aussi rapidement que les chercheurs le souhaitent.
- c) La connaissance et la perception que nous avons de la communauté de la recherche en ce qui concerne la confidentialité ne sont pas suffisamment claires pour instaurer la confiance. Nous avons l'impression qu'un accès aisé aux données risque de nuire à la confiance de nos fournisseurs et, par conséquent, à la qualité des données de sortie et à l'image d'un institut de statistique.
- d) Les méthodes, instruments et compétences dont nous disposons ne sont pas suffisants pour nous permettre d'évaluer les risques de divulgation et de suivre les utilisateurs de données confidentielles.

### 3. Le rôle des concepteurs de méthodes et d'instruments pour le CDS

De nombreux efforts ont été accomplis au cours des dernières décennies pour élaborer des méthodes et développer des outils logiciels pour le CDS.

Cela n'est pas suffisant et les INS ne se sentent pas en sécurité lorsqu'ils publient les données dans le cadre méthodologique existant.

Des travaux complémentaires ainsi qu'une collaboration entre la communauté des chercheurs et les statisticiens nationaux sont nécessaires pour améliorer le cadre conceptuel et l'évaluation de la confidentialité statistique.

Le projet financé dans le cadre du 5<sup>ème</sup> PC permettra sans aucun doute d'obtenir de précieux résultats; il est néanmoins nécessaire de préparer dans les prochains mois un nouveau projet qui s'inscrira dans le cadre du 6<sup>ème</sup> PC.

Une telle activité ne peut se limiter à l'Europe, celle-ci doit donc être le projet commun d'un réseau d'excellence au niveau mondial.

### 4. Le rôle des INS

Les INS sont en première ligne en ce qui concerne le traitement de la confidentialité statistique; leur mission comporte de nombreuses facettes:

- informer les fournisseurs de données;
- organiser le traitement des données confidentielles;
- traiter les données confidentielles;
- utiliser les données confidentielles pour produire des statistiques destinées à être diffusées;
- transmettre les données confidentielles (à Eurostat);
- accéder souvent à des données confidentielles à des fins de recherche;
- gérer les questions juridiques, administratives et techniques.

Lorsque l'on considère les nouveaux besoins des pays candidats ainsi que de nombreux autres pays coopérant avec la communauté statistique internationale, on peut imaginer l'importance et le volume de la tâche, surtout lorsqu'il s'agit de répondre aux besoins de formation.

### 5. Le rôle du CEIES

Il est important pour nous que le CEIES soit parfaitement conscient de l'importance du sujet et que les actions déjà mises en œuvre apportent leur contribution au prochain pas significatif que nous allons faire dans ce domaine.

Il est important que le CEIES adresse un message fort aux décideurs politiques au niveau européen et national afin:

1. d'améliorer la dimension juridique et de renforcer le besoin de dispositions juridiques appropriées qui permettront un accès plus aisé à des fins de recherche.
2. de souligner la nécessité de recevoir des ressources adéquates dans le cadre des programmes européens (R&D, e-Europe) ainsi qu'au niveau national, en se concentrant sur cette question de la confidentialité statistique et de l'accès aux microdonnées à des fins de recherche.

## 6. Eurostat

En ce qui concerne le plan d'action d'Eurostat, nous tenterons de faire de notre mieux avec les ressources dont nous disposons. D'autres priorités porteront sur:

- la mise en œuvre du règlement n° 831/2002;  
l'établissement des dispositions, du cadre administratif et des modalités techniques nécessaires à l'octroi d'autorisations;  
la création d'un "centre sécurisé" à Eurostat;
- le renforcement de la coopération avec les INS des EM afin d'établir ensemble un plan d'action pour les trois prochaines années;
- le renforcement de la coopération internationale par la création d'un réseau de collaboration afin de progresser dans le domaine de la confidentialité statistique.

Ces actions comprendront des contacts avec les fournisseurs et utilisateurs de données, la communauté de la recherche et les autorités légales.

## 7. Communication

La question de la confidentialité statistique est bien connue des statisticiens officiels, relativement bien connue des chercheurs (essentiellement ceux qui ont été confrontés au problème de l'accès), très peu connue des fournisseurs, c'est-à-dire des sujets concernés (nous devons mieux comprendre leur perception) et probablement très peu connue des personnes chargées de la protection des données et des politiciens.

Il est certain que les connaissances et les préoccupations au sujet de la confidentialité augmentent probablement aussi rapidement que la nécessité pour les chercheurs d'avoir accès aux données confidentielles.

Dans cet examen de la question de l'accès, nous, statisticiens et chercheurs, ne devons pas rester seuls, chacun de notre côté, nous faisant face avec sympathie ou agressivité. Nous devons impliquer dans ce processus les autres partenaires, les fournisseurs de données, les autorités légales et politiques.

En Europe, nous n'avons ni "Comité européen pour l'éthique et la confidentialité", ni "Association statistique européenne", ni "Association européenne de la recherche".

Sans implication des chercheurs dans le dialogue social, il est difficile de progresser.

## SOLUTIONS INNOVANTES POUR PERMETTRE UN ACCÈS AUX MICRODONNÉES: RÉCAPITULATIF

**Patrick T. Geary**

Les documents présentés au cours du séminaire et le débat auquel ils ont donné lieu ont été de grande qualité. Bon nombre des questions soulevées lors du séminaire ayant déjà été parfaitement résumées par Photis Nanopolous dans sa contribution, je propose de limiter mes commentaires à deux points précis.

En premier lieu, il est à noter que la question de la confidentialité revêt une importance centrale dans toute discussion sur la fourniture d'un accès aux microdonnées. La protection de l'identité des fichiers de données individuelles est indispensable. Il a été souligné que la suppression des identificateurs individuels de base n'était pas suffisante pour garantir l'anonymat des données. Différents moyens d'atteindre cet objectif ont été présentés durant le séminaire, et notamment les techniques de "masquage des données", destinées à contrecarrer les activités des "fouineurs" (*data snoopers*). L'imposition d'obligations contractuelles aux personnes et institutions recevant des microdonnées a également été envisagée. Les forces et faiblesses respectives des diverses méthodes ont été discutées en détail, étant entendu qu'une certaine anonymisation des données constituait une condition préalable à leur diffusion. La possibilité qu'une technique particulière de masquage des données, tout en garantissant l'anonymat, s'avère efficace pour certaines applications des microdonnées, mais provoque des distorsions pour d'autres applications des mêmes données a été évoquée. Bien que l'accent ait été mis sur le fait qu'il n'existait pas de preuves démontrant que des chercheurs avaient violé des accords contractuels visant à protéger la confidentialité des données, plusieurs cas ont néanmoins été cités dans lesquels la publication de résultats par inadvertance avait permis une identification individuelle.

Le débat sur la confidentialité a concerné exclusivement la diffusion de microdonnées par un institut national de statistique ou une autre autorité publique à des chercheurs. Le contexte antérieur de la fourniture de données **entièrement identifiées** par les répondants individuels aux instituts nationaux de statistique et aux organismes publics n'a pas été mentionné. En principe, le risque de divulgation de données individuelles par les membres du personnel d'un institut national de statistique ou d'un organisme public est plus élevé, car certains d'entre eux ont accès à l'ensemble des détails des fichiers individuels. Toutefois, leur obligation légale de protéger l'anonymat et leur réputation acquise au fil des ans sont, dans l'opinion publique, largement considérées comme des garanties suffisantes. Bref, le public leur fait confiance. Il est cependant possible de faire valoir que l'accès des chercheurs à l'ensemble des détails des fichiers individuels n'est pas un problème, vu que les coûts encourus par eux-mêmes et leurs employeurs en cas de violation des obligations contractuelles visant à protéger les données individuelles sont énormes et que l'expérience acquise à ce jour en matière de diffusion de microdonnées aux chercheurs permet d'accorder à ces derniers une confiance comparable.

En second lieu, il convient de s'interroger sur les raisons d'une telle croissance de la demande d'accès aux microdonnées. Faut-il y voir simplement l'expression d'un désir d'effectuer des recherches scientifiques à partir d'ensembles de données riches? La réponse à cette question est clairement non. Bon nombre de recherches importantes dans les domaines social et économique, qu'elles présentent un caractère fondamental ou un intérêt pour la formulation et l'évaluation des politiques publiques, peuvent **uniquement** être entreprises à l'aide de microdonnées; elles ne peuvent pas être menées en exploitant des statistiques publiées ou des données agrégées. Les documents présentés lors de la troisième session du séminaire illustrent parfaitement ce dernier point. Il faut toutefois préciser qu'ils ne représentent qu'un petit sous-échantillon des questions dont l'étude **requiert** l'accès à des microdonnées. D'importants motifs d'intérêt public peuvent être invoqués pour accorder un accès aussi large que possible, dans le respect de prescriptions raisonnables en matière de confidentialité.

Si les instituts nationaux de statistique ne prennent pas eux-mêmes les mesures nécessaires pour répondre à cette demande, ils risquent de se faire court-circuiter et les décideurs politiques s'adresseront alors à des instituts de recherche, des universités, voire des sources commerciales, pour se procurer les informations requises. Dès à présent, un grand nombre de travaux de recherche utilisent des microdonnées ne provenant pas des instituts nationaux de statistique. À l'occasion du séminaire ont été évoqués l'exemple du panel socio-économique allemand, ainsi que la nouvelle enquête sociale européenne, une initiative trouvant son origine dans le besoin de davantage de comparaisons à l'échelle européenne.

Des progrès considérables ont incontestablement été accomplis dans le domaine de l'accès aux microdonnées, mais leur répartition n'est pas uniforme: au sein de l'UE, certains pays ont progressé bien plus que d'autres. Le règlement 831/2002 de la Commission (Eurostat) marque cependant un important pas en avant, à mettre au crédit de tous les acteurs concernés. Eurostat et les instituts nationaux de statistique sont conscients des conséquences en termes de ressources de la fourniture d'un accès élargi aux microdonnées. Dans le cadre de la poursuite de cet objectif, ils ont droit à tout le soutien que le CEIES pourra obtenir aux niveaux tant de l'UE que des États membres.

## LISTE DES PARTICIPANTS

AGRESTI Anna Maria  
Research Analyst  
European Central Bank  
Keiserstrasse, 25  
Frankfurt  
GERMANY  
Anna\_maria.agresti@ecb.int

AINARS Uldis  
Director, Information, Publishing  
& Poligraphy Department,  
1, Lacplesa Str.  
LV-1301 Riga  
Uainars@csb.lv

ALBERGARIA Henrique  
INE Portugal  
Coimbra  
PORTUGAL  
Henrique.Albergaria@ine.pt

ANDERSEN Otto  
Statistics Denmark  
Sejrogade 11  
2100 Copenhagen Ø  
DENMARK  
Oan@dst.dk

ARNAUDOVA Ekaterina  
Vice-president of National  
Statistical Institute of Bulgaria  
2, P. Volov Str.  
Sofia  
BULGARIA  
Earnaudova@nsi.bg

BACELAR Mr S.  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL

BARBIERI Giovanni  
Director  
Dissemination  
ISTAT  
Via Torino 6  
ITALY  
Barbieri@istat.it

BARCENA MARTIN Elena  
Malaga University  
Statistics Department  
Campus El Egido s/n 290  
E-SPAIN  
barcena@uma.es

BARNES Matthew  
Economist  
Business Data Linking Branch  
Zone D4-16,  
1 Drummond Gate,  
London, SW1V 2QQ  
UNITED-KINGDOM  
matthew.barnes@ons.gov.uk

BARTOLONI Rita  
Dirigente Ufficio Istat  
Via Cesare Balbo, 1  
Perugia,  
ITALY  
Ribartol@istat.it

BENTO Maria da Graça  
Conselho Superior de Estatística  
Avenida António José de Almeida  
Lisboa  
PORTUGAL  
Mgraca.bento@ine.pt

BLUNDELL Richard  
Professor of Economics  
Gower Street  
London WC1 E6BT  
UNITED KINGDOM  
r.blundell@ucl.ac.uk

BOYKO Ernie  
Director, Library & Information Centre  
Statistics Canada  
2-0 RH Coats Building  
Ottawa (Ontario)  
CANADA  
Eboyko@statcan.ca

BURGET Dieter  
Statistics Austria  
Hintere Zollamtsstr. 2b  
Wien  
AUSTRIA  
Dieter.burget@statistik.gv.at

BYK Daniel  
European Commission  
Director, Eurostat  
BECH A4/004  
Jean Monnet Building  
L-2920 Luxembourg  
Daniel.byk@cec.eu.int

CALIÇO Orlando  
Director  
Banco de Portugal  
Av. Almirante Reis No. 71  
Lisboa  
PORTUGAL  
Ocalico@bportugal.pt

CARIANI Giovanni  
Istituto Nazionale di Statistica  
Via Depretis, 77 00184  
Roma ITALIA  
Cariani@istat.it

CARICCHIA Alfonsina  
Direttore Centrale  
ISTAT  
Via Depretis, 77 00184  
Roma ITALIA  
Caricchi@istat.it

CHARLIER Hubert  
European Commission  
ESTAT  
BECH B2/472  
Batiment Jean Monnet  
L-2920 Luxembourg  
hubert.charlier@cec.eu.int

CHATZIDOUKAKIS Efstratios  
European Commission  
Legal Adviser, Eurostat  
BECH B4/389  
Jean Monnet Building  
L-2920 Luxembourg  
Efstratios.chatzidoukakis@cec.eu.int

CLARK Alex  
Office for National Statistics  
Segensworth Road  
Titchfield  
Fareham Hants  
UNITED KINGDOM  
Alex.clark@ons.gov.uk

CLODE Teresa  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL

COSTA Maria João  
GPLP do MJ  
Av. Óscar Monteiro Torres – 39  
Lisboa  
PORTUGAL  
MjoaoCosta@gplp.mj.pt

COX Lawrence  
Associate Director  
National Center for Health Statistics  
6525 Belcrest Road, Room 915  
Hyattsville, MD20782, U.S.A.  
LCOX@CDC.GOV

CUNHA Catarina  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL  
Catarina.cunha@ine.pt

DALE Angela  
Director, Cathie Marsh Centre  
for Census and Survey Research  
Faculty of Social Studies and Law  
University of Manchester  
Manchester M13 9PL  
UNITED KINGDOM  
Angela.dale@man.ac.uk

DEBUSSCHERE Marc  
Statistics Belgium  
Leuvenseweg 44  
B – 1000 BRUSSELS  
Marc.debusschere@statbel.fgov.be

DIOGO Susana  
Planning and Study Dept. of the Ministry  
& Public Works, Housing &  
Transport  
Av. Columbano Bordalo Pinheiro, 5-2º  
1070 – 060 Lisboa  
PORTUGAL  
Sdiogo@gepmes.pt

DOLENC Danilo  
Statistical Office of the Republic of  
Slovenia  
Parmova 33  
Ljubljana  
SLOVENIA  
Danilo.dolenc@gov.si

DUARTE Manuel João  
DETEFP-MSST  
Rua Rodrigo da Fonseca, 55  
Lisboa PORTUGAL  
Mjoao@detefp.pt

DUARTE DE CARVALHO Filipa  
Instituto Superior de Economia e Gestão  
Rua do Quelhas 6  
1200 Lisboa  
PORTUGAL  
Filipadc@iseg.utl.pt

DUNCAN George  
Professor of Statistics/Associate Dean for  
Faculty  
Carnegie Mellon University  
5000 Forbes Ave  
Pittsburgh PA 15217  
U.S.A.  
Gd17@andrew.cmu.edu

EPLER Margit  
Chamber of Labour  
Prinz-Eugene str. 20-22  
A – 1040 Vienna  
Margit.epler@akwien.or.at

ERIKSON Robert  
Professor  
Swedish Council for Working Life &  
Social Research  
Box 2220  
SE – 10315 Stockholm  
Robert.erikson@fas.forskning.se

ESTEVEENS Luzia  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL  
Luzia.estevens@ine.pt

EVANS Deborah  
European Commission  
Eurostat  
CEIES Secretariat  
BECH A4/121  
Jean Monnet Building  
L-2920 Luxembourg  
Debora.evans@cec.eu.int

FRANCO Ana  
European Commission  
ESTAT  
BECH B4/433  
Batiment Jean Monnet  
L-2920 Luxembourg  
Ana.franco@cec.eu.int

GEARY Patrick  
Economics Department  
National University of Ireland, Maynooth  
Maynooth  
Co. Kildare, Ireland  
Patrick.Geary@may.ie

GLAVIE Christian  
National Bank of Belgium  
Head of Dissemination Section  
Bvd de Berlaimont 14  
B- 1000 Brussels  
Christian.glavie@nbb.be

GOMES Paulo  
Presidente  
INE Portugal  
Av. António José de Almeida  
1000-043 Lisboa  
Paulo.gomes@ine.pt

HELDAL Johan  
Statistics Norway  
P.O. Box 8131, Dep  
Oslo  
NORWAY  
Johan.heldal@ssb.no

HÖHNE Jörg  
Senior Researcher  
Berlin State Statistical Office  
Alt-Friedrichsfelde 60  
D-10315 Berlin  
J.Hoehne@statistik-berlin.DE

ILARIONOV Ilarion  
Vice-president of  
National Statistical Institute of Bulgaria  
2, P. Volov Str.  
Sofia  
BULGARIA  
Iilarionov@nsi.bg

IMBRIANI Cesare  
Commissione  
Garanzia Informazione Statistica  
Via Po 16/A  
00198 Roma  
ITALY  
Imbriani@uniroma1.it

JACKSON Paul  
Principle Methodologist  
Office for National Statistics  
Segensworth Road  
Fareham, Hampshire  
UNITED KINGDOM  
Paul.j.jackson@ons.gsi.gov.uk

JEDOMSKIHH Natalja  
Statistical Office of Estonia  
15 Endla Str  
Tallinn  
ESTONIA  
Natalja.jedomskihh@stat.ee

KING John  
European Commission  
Eurostat  
BECH A2/166  
Jean Monnet Building  
L-2920 Luxembourg  
John.king@cec.eu.int

KOCH Iris  
Bundesanstalt für Arbeit  
Postfach  
D-90327 NÜRNBERG  
Germany  
iris.koch@arbeitsamt.de

KOTULICS Tamas  
Hungarian Central Statistical Office  
P.O. Box 51  
H – 1525 Budapest  
Tamas.kotulics@ksh.gov.hu

KRAUS Franz  
Head of Eurodata Research Archive  
Mannheim Centre for European Social  
Research  
L7, 1  
Mannheim  
GERMANY  
Franz.kraus@mzes.uni-mannheim.de

LACOMBA ARIAS Beatriz  
Malaga University  
Statistics Department  
Campus El Egido s/n 290  
E-SPAIN  
beatriz@uma.es

LAMEL Joachim  
Untere Oeden 28  
A-3400 Klosterneuburg  
Joachim.lamel@wko.at

LANZIERI Giampaolo  
European Commission  
Eurostat  
BECH B2/457  
Jean Monnet Building  
L-2920 Luxembourg  
Giampaolo.lanzieri@cec.eu.int

LAUWERIJS Nicole  
European Commission  
Eurostat  
CEIES Secretariat  
BECH A4/123  
Jean Monnet Building  
L-2920 Luxembourg  
Nicole.lauwerijs@cec.eu.int

LEWIS David  
8 rue de l'Ernz  
Eisenborn  
LUXEMBOURG  
Dailewis@pt.lu

LO MORO Vincenzo  
ISTAT  
Via Cesare Balbo 16  
Roma  
ITALY  
Lomoro@istat.it

MACLEAN Ian  
Statistics Users' Council  
Lancaster House - More Lane  
UK-Esher, Surrey KT108AP  
Bts@dial.pipex.com

MADELENO Margarida  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL

MAHRINGER Helmut  
Economist,  
Austrian Institute of Economic Research,  
PO Box 91  
A-1103 Vienna  
Mahring@wifo.ac.at

MALIRANTA Mika  
Head of Research,  
Statistics Finland  
Box 7  
00022 Helsinki  
FINLAND  
Mika.maliranta@stat.fi

MARCHI Marco  
Professor of Health Statistics  
University of Florence  
Dept. Statistics-Viale Morgagni 59  
50134 Firenze  
ITALY  
Marchi@ds.unifi.it

MARKELEVICIUS Jonas  
Deputy Director General  
Gedimino Ave, 29  
Vilnius  
LITHUANIA  
Statistika@mail.std.lt

MATOS Ana  
Instituto Nacional de Estatística  
Avenida António José de Almeida  
Lisboa  
PORTUGAL  
Ana.matos@ine.pt

McCAA Robert  
Professor  
Minnesota Population Center  
537 Heller Hall  
University of Minnesota  
Minneapolis, MN  
U.S.A.  
Rmccaa@umn.edu

MEINKEN Holger  
Bundesanstalt für Arbeit  
Postfach  
D-90327 NÜRNBERG  
Germany  
holger.meinken@arbeitsamt.de

MENDES Maria de Luz  
GPPAA  
Rua Padre António Vieira  
1099-073 Lisboa  
PORTUGAL  
Luzmendes@gppaa.min-agricultura.pt

NANOPOULOS Photis  
European Commission  
Director, Eurostat  
BECH A4/172  
Jean Monnet Building  
L-2920 Luxembourg  
Photis.nanopoulos@cec.eu.int

NÄSLUND-FOGELBERG Annika  
European Commission  
Eurostat  
CEIES Secretariat  
BECH A4/124  
Jean Monnet Building  
L-2920 Luxembourg  
Annika.naslund-fogelberg@cec.eu.int

NEDELJKOVIC Ranko  
Deputy Director  
Federal Statistical Office  
Kneza Milosa, 20  
Belgrade  
YUGOSLAVIA  
Ranko@szs.sv.gov.yu

NOBEL Joris  
Statistics Netherlands  
P.O. Box 4000  
2270 JM Voorburg  
THE NETHERLANDS  
JNBL@CBS.nl

NOČ RAZINGER Mojca  
Director Counsellor  
Statistical Office of the Republic of  
Slovenia  
Vozarski pot 12  
Ljubljana  
SLOVENIA  
Mojca.noc@gov.si

PERPÉTUO Fernanda  
INE Portugal  
Rua Padre Luís Aparício 5  
1150-248 Lisboa  
PORTUGAL  
Fernanda.perpetuo@ine.pt

PETTERSSON Birgitta  
Legal Adviser  
Statistics Sweden  
Box 24300  
Stockholm  
Sweden  
Birgitta.pettersson@scb.se

REBELO Maria  
DETEFP-MSST  
R. Rodrigo da Fonseca, 55  
Lisboa  
PORTUGAL  
Mrebello@detefp.pt

SCHIMPL-NEIMANNS Bernhard  
Dept. Superintendent of microdata  
Zentrum für Umfragen, Methoden und  
Analysen  
Postfach 12 21 55  
D-68072 Mannheim  
Schimpl-neimanns@zuma-mannheim.de

SCHÖBERL Marianne  
Austrian Institute of Economic Research  
PO Box 91  
A-1103 Vienna  
Marianne.schoeberl@wifo.ac.at

SCHREY Ernst  
Head of Division  
Statistisches Bundesamt-II D  
Gustav-Stresemann-Ring 11  
D – 65189 Wiesbaden  
Ernst.schrey@destatis.de

SCHÜRER Kevin  
Director  
UK Data Archive  
University of Essex  
Wivenhode Park  
Colchester, Essex  
UNITED KINGDOM  
Schurer@essex.ac.uk

SERI Giovanni  
ISTAT  
Via Cesare Balbo, 16  
Rome  
ITALY  
Seri@istat.it

SESTITO Paolo  
Ministero del Lavoro e delle Politiche  
Sociali  
Via Fornovo 8, Building B, 4th Floor  
ROMA, ITALY  
Psestito@minwelfare.it

SIUNE Karen  
Director  
Analyseinstitut for Forskning  
Finlandsgade 4  
DK- 8200 Århus N  
Ks@afsk.au.dk

SKJÅK Knut Kalgraff  
Assistant Director  
Norwegian Social Science Data Services  
Hans Holmboesgt 22  
5007 Bergen  
NORWAY  
Skjak@nsd.uib.no

SPIEKER Finn  
Statistics Denmark  
Sejrogade 11  
2100 Copenhagen  
DENMARK  
Fsp@dst.dk

STEVENS Luzia  
Instituto Nacional de Estatística  
Av. Antonio José de Almeida  
Lisboa  
PORTUGAL

STOOP Ineke  
Social and Cultural Planning Office  
P.O. Box 16164  
2500 BD The Hague  
THE NETHERLANDS  
Stoop@scp.nl

STURM Roland  
Federal Statistical Office  
Gustav Stresemann Ring 11  
D-65189 Wiesbaden  
Roland.sturm@destatis.de

SUOMINEN Risto  
Federation of Finnish Enterprises  
Kaisaniemenkatu 13 A  
00100 Helsinki  
FINLAND  
Risto.suominen@suomen.yrittajat.fi

TEMPLETON Miles  
Agricultural Census Data Manager  
DEFRA  
Foss House, Peasholme Green  
York, U.K.  
Miles.h.templeton@defra.gsi.gov.uk

TORRES BORELL Gloria  
Instituto Nacional de Estadística  
Capitán Haya, 51  
28071 Madrid  
SPAIN  
Gtorres@ine.es

TRIVELLATO Ugo  
Universita di Padova –  
Dip. Scienze Statistiche  
Via Cesare Battisti, 241  
35121 Padova  
ITALY  
Trivell@stat.unipd.it

VALEN Thorleiv  
Statistics Norway  
P.O.B. 8131 Dep  
N-0033 Oslo  
Thorleiv.valen@ssb.no

VAN DONINCK Bogdan  
Federal Office for Scientific,  
Technical and Cultural Affairs  
Wetenschapsstraat 8  
B-1000 BRUSSEL  
Vdon@belspo.be

VEIGA Conceição  
Instituto Nacional de Estatística  
Av. António José de Almeida  
Lisboa  
PORTUGAL

VICENTE Maria João  
ISCTE  
Rua Jose Carlos de Melo no. 26  
Almada  
PORTUGAL  
Mjvicente@mail.telepac.pt

VORGRIMLER Daniel  
Federal Statistical Office  
Gustav-Stresemann Ring 11  
D-65189 Wiesbaden  
Daniel.vorgrimler@destatis.de

WAGNER Gert  
TUB and DIW Berlin  
Koenigin Luise Strasse 5  
D- 14195 Berlin  
Gwagner@diw.de

WATKINS Wendy  
Co-ordinator, Data Centre  
Carleton University  
1125 Colonel By Drive  
Ottawa  
CANADA  
Wwatkins@ccs.carleton.ca

WESTERGAARD-NIELSEN Niels  
Aarhus School of Business  
4 Fuglesangs Allé  
DK – 8210 Aarhus  
Nwn@cls.dk

ZWICK Markus  
Federal Statistical Office  
Gustav-Stresemann-Ring 11  
D - 65180 Wiesbaden  
Markus.zwick@destatis.de