

No 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2005 : Défis
méthodologiques reliés aux
besoins futurs d'information**



2005



Statistique
Canada

Statistics
Canada

Canada

CONTRÔLE DE LA DIVULGATION STATISTIQUE : CADRE JURIDIQUE ET ASPECTS MÉTHODOLOGIQUES

Maria João Santos¹

RÉSUMÉ

Tous les pays membres de l'Union européenne (UE) font face à des problèmes similaires de contrôle de la divulgation statistique (CDS). Tous doivent trouver un juste équilibre entre la protection des renseignements confidentiels fournis par les répondants et les demandes fort légitimes de la société, des chercheurs et des décideurs qui souhaitent obtenir des données de plus en plus détaillées. Cette demande croissante suscitée par l'avènement de l'ère de l'information et de la société du savoir est un problème commun du Système statistique européen (SSE). Le CDS représente également un problème fondamental pour Eurostat, parce qu'il est l'élément central de la confiance fragile que les fournisseurs de données ont dans les organismes recueillant des statistiques. Il influence profondément la qualité des statistiques de l'Union européenne et, par conséquent, la relation entre Eurostat et le SSE. En outre, le cadre réglementaire dans lequel sont produites les statistiques comprend des règles strictes visant à assurer que l'information fournie par les répondants est protégée comme il se doit contre la divulgation. Afin de pouvoir relever les défis qui se posent à l'échelle européenne, les problèmes de CDS connexes devront être abordés par tous les pays dans les années à venir. Dans cet article on expose les enjeux et les stratégies relatives à la confidentialité actuellement en vigueur à Eurostat.

1. INTRODUCTION

Cet article vise à fournir un aperçu des divers enjeux liés à la confidentialité dans une perspective européenne. Elle vise à donner aux experts techniques une idée des difficultés qui se posent dans le contexte multinational et administratif, et qui pourraient ne pas être perceptibles à prime abord. La gamme variée de perceptions et l'absence d'une norme bien définie sont une source de diversité qui aggrave beaucoup les problèmes courants de confidentialité au niveau européen. Dans cet article on préconise un partenariat étroit entre les collectivités des administrateurs et des chercheurs, ainsi que des interventions et des responsabilités bien définies dans le domaine de la recherche scientifique, en vue de concevoir des pratiques éprouvées à l'appui de la réflexion au niveau juridique en Europe.

2. CADRE JURIDIQUE DE LA CONFIDENTIALITÉ

2.1 Cadre général

Le droit à la protection de la vie privée est un droit fondamental. Il comprend la protection des personnes dans le contexte du traitement des données personnelles. Cela signifie, par exemple, le droit de recevoir certaines données, le droit d'accéder aux données, le droit de faire corriger les données, etc. La confidentialité statistique vise principalement la protection de la vie privée dans le domaine statistique, et joue un rôle clé à l'égard de la confiance nécessaire qui doit exister entre les organismes statistiques et les répondants. La fiabilité et la précision des données de base reposent sur la confiance mutuelle et cette dernière permet en dernier ressort la production de statistiques de grande qualité.

Au niveau de l'Union européenne (UE), la confidentialité statistique est abordée dans les actes juridiques suivants :

- Règlement (Euratom, CEE) n° 1588/90 du Conseil, du 11 juin 1990, relatif à la transmission à l'Office statistique des Communautés européennes d'informations statistiques couvertes par le secret;
- Règlement (CE) n° 322/97 du Conseil du 17 février 1997 relatif à la statistique sur les communautés;

¹ Unité de la méthodologie et de la recherche, Eurostat, L-2920, Luxembourg

- 97/281/CE : Décision de la Commission du 21 avril 1997 concernant le rôle d'Eurostat en matière de production de statistiques sur les communautés;
- Règlement (CE) n° 831/2002 de la Commission du 17 mai 2002 portant modalité d'application du règlement (CE) n° 322/97 du Conseil relatif à la statistique sur les communautés en ce qui concerne l'accès aux données confidentielles à des fins scientifiques;
- 2004/452/CE : Décision de la Commission du 29 avril 2004 établissant la liste des organismes dont les chercheurs peuvent être autorisés à accéder à des données confidentielles à des fins scientifiques.

La confidentialité statistique est réglementée au niveau de l'UE uniquement en ce qui a trait aux activités statistiques menées par Eurostat et les responsables statistiques nationaux pour la production de statistiques communautaires. Des modalités particulières en matière de confidentialité coexistent encore au niveau national, et elles peuvent différer des modalités de confidentialité statistique de l'UE. Ces différences ont moins trait à la substance (les concepts généraux sont identiques dans une très large mesure) qu'à la perception de l'enjeu (le cadre national demeure le cadre de référence), ce qui est tout aussi important.

De ce fait, les modalités de confidentialité de statistiques existantes ne font pas l'objet d'un règlement uniforme, ce qui entraîne des difficultés d'interprétation entre les États membres et la Commission, et complique les travaux dans différents secteurs. L'amélioration du cadre existant devrait contribuer à éviter des discussions répétées et, dans certains cas, à surmonter des obstacles lorsqu'il s'agit de traiter les enjeux de confidentialité dans le contexte de la négociation de règlements sectoriels.

À l'heure actuelle, on réfléchit toujours à Eurostat et au niveau des États membres sur la nécessité de réviser le cadre juridique, selon les principes de maximisation de la qualité des statistiques européennes produites à la fois par les États membres et par les institutions européennes, ainsi que d'augmenter la possibilité d'une utilisation secondaire des données par les collectivités de chercheurs et le grand public, tout en respectant le mandat de confidentialité, qui consiste à préserver la divulgation directe ou indirecte de données individuelles.

La révision proposée pourrait faire l'objet de modifications du cadre juridique dans plusieurs domaines. En ce qui a trait à la transmission des données confidentielles, on pourrait envisager la modification des dispositions figurant dans l'article 14 du Règlement 322/97 sur la transmission de données confidentielles ne permettant pas une identification directe, en vue de les remplacer par des modalités selon lesquelles la transmission et l'échange devraient couvrir les données confidentielles définies objectivement selon l'article 13 du même règlement, et d'englober ainsi toute la gamme de données confidentielles. Ces transmissions et échanges devraient être permis : entre les États membres et entre les États membres et Eurostat, dans la mesure où cela a trait et est nécessaire à la production et à la qualité des statistiques sur les communautés.

Le concept selon lequel les données tirées de sources accessibles au public ne devraient pas être considérées comme confidentielles, qui est déjà abordé dans l'article 13 du Règlement n° 322/97, devrait être appliqué de façon plus systématique, peut-être par l'entremise d'un acte juridique particulier définissant les variables et les domaines disponibles dans le public, selon les directives pertinentes de l'UE. En parallèle, l'article 13 pourrait être modifié, afin de faciliter sa mise en œuvre, grâce au retrait de la spécification suivante : « et qui restent accessibles à celui-ci (le public) auprès des autorités nationales », qui est perçue comme une contrainte additionnelle à l'égard de la mise en œuvre.

L'acceptation large d'une base objective pour la déclaration des données confidentielles et la mesure du risque de divulgation faciliteraient définitivement les progrès juridiques dans le domaine de la confidentialité statistique. Il faut de toute évidence que les chercheurs scientifiques exercent leur pouvoir, afin de mettre un terme à cette discussion subjective sans fin. Les avocats attendent des conseils techniques avisés avant de concevoir des dispositions législatives harmonisées.

2.2 Accès pour les chercheurs

On se rend de plus en plus compte des avantages que comporte l'accès aux microdonnées pour la recherche et l'analyse. Parallèlement, il est essentiel d'assurer la protection de la confidentialité des données. Il faut que de nouvelles approches soient élaborées, en vue de répondre à ces objectifs qui exercent des pressions conflictuelles. Les risques pour la confidentialité doivent être gérés efficacement. L'un des défis clés consiste à déterminer la façon de réduire les risques pour la confidentialité, y compris la perception des menaces. Il est essentiel d'établir un juste équilibre.

L'élaboration de politiques complexes nécessite une réflexion causale fondée sur des variables multiples concernant les options stratégiques qui, quant à elles, reposent sur des données complexes, multidimensionnelles et souvent longitudinales. Au fur et à mesure que l'économie se complexifie, et que la population se diversifie, des données et une analyse de données de plus en plus détaillées sont nécessaires pour que les politiques correspondent bien aux options économiques et démographiques.

Un partenariat public-privé efficace entre les organismes de collecte de données et la collectivité des chercheurs constitue un élément essentiel de la prise en compte des analyses de données complexes, et plus particulièrement de microdonnées, dans la conception et l'évaluation des politiques. Ce partenariat entre les INS et les chercheurs profite aux deux parties et bénéficie de l'amélioration continue de l'accès aux données, à la fois par l'entremise d'ensembles de données à grande diffusion et de modalités restreintes d'accès aux données. Le rapport entre l'utilisation et la qualité des données représente la base essentielle de l'intérêt commun du système statistique et de la grande collectivité des chercheurs à l'égard d'un accès large et responsable aux données.

Il est nécessaire d'explorer de nouvelles avenues pour donner aux chercheurs l'accès aux données et améliorer en parallèle les instruments actuels.

Rationalisation de l'application du Règlement de la Commission n° 831/2002

Une description et une analyse détaillées de cet acte juridique figurent dans la communication présentée par John King et Jean Louis Mercy, dans le cadre de la séance de travail sur la confidentialité des données statistiques tenue au Luxembourg, du 7 au 9 avril 2003. Même si ce règlement suscite des espoirs importants en ce qui a trait à la disponibilité de microdonnées pour la collectivité des chercheurs, son application a causé plusieurs difficultés, qui ont fait en sorte qu'elle a progressé lentement.

Dans sa réunion de décembre 2004, le Comité du secret statistique (CSS) a analysé les progrès liés à l'application de ce règlement et a approuvé l'élaboration de procédures relatives au traitement rapide des demandes des chercheurs et à l'admissibilité des institutions de recherche. Ces procédures accélérées seront soumises au CSS, à sa prochaine réunion, en décembre 2005; leur adoption améliorera l'actualité et l'efficacité du règlement.

Il existe deux niveaux d'accès aux microdonnées :

Niveau un : **Données confidentielles obtenues des autorités nationales.** Seule l'identification indirecte des unités statistiques concernées est permise. Cet accès se fait par l'entremise d'un centre sécuritaire à Eurostat.

Niveau deux : **Ensembles de microdonnées anonymes extraites des données mentionnées précédemment.** Il s'agit d'enregistrements statistiques individuels qui ont été modifiés, afin de réduire le risque d'identification des unités statistiques auxquelles ils se rapportent, conformément aux pratiques éprouvées en vigueur actuellement.

Cet accès se fait grâce à la distribution de CD-ROM chiffrés, conformément aux contrats conclus entre Eurostat et les institutions concernées.

À l'heure actuelle, les microdonnées de niveau deux destinées aux chercheurs peuvent être fournies uniquement pour trois domaines statistiques. Il s'agit du Panel Communautaire des Ménages (PCM), de l'Enquête sur la formation professionnelle continue (EFPC) et de l'Enquête sur la population active (EPA). Par ailleurs, le Groupe de travail de l'enquête communautaire sur l'innovation examine actuellement des critères en vue de la distribution des fichiers de microdonnées de cette enquête. Par ailleurs, un groupe de travail a été établi pour faire la même chose relativement à l'Enquête européenne sur le revenu et les conditions de vie (EU-SILC).

Des mesures devront être prises pour proposer l'ajout d'autres ensembles de microdonnées à ceux mentionnés dans le Règlement n° 831/2002 de la Commission, comme l'ESG (Enquête sur la structure des gains).

Les possibilités offertes dans le cadre de ce règlement ont pour avantage que les chercheurs ont maintenant la possibilité d'accéder à des ensembles de données harmonisées qui couvrent tous les États membres. Auparavant, l'accès aux données de chacun des EM nécessitait un long processus de demande auprès de chacun. Cela fournit aux chercheurs la possibilité d'effectuer des recherches et des analyses à l'échelle de l'Union européenne. Le tableau ci-dessous présente une synthèse des projets déclarés par les établissements de recherche qui, en 2004, ont soumis à Eurostat des demandes de microdonnées du Panel Communautaire des Ménages (PCM).

Contrats de recherche à partir des données du PCM, année 2004, principaux sujets	
<i>Études de sous-populations particulières</i>	<i>Études de phénomènes particuliers</i>
Personnes âgées	Mobilité
Personnes pauvres	Inégalité du revenu
Régions	Transition de l'emploi au chômage
Chômage à long terme	Imposition, subventions
Femmes mariées	Transferts à l'intérieur des familles
Participation des femmes au marché du travail	Inégalité dans le revenu et la scolarité
Personnes divorcées	Modifications de la rémunération
Travailleurs temporaires	Éducation et santé
Personnes à la fin de leur vie active	Participation au marché du travail et fécondité
Jeunes	Soins des enfants
	Discrimination

Le Règlement n° 831/2002 laisse entrevoir (article 3) un processus de demande relativement simple pour les chercheurs appartenant à deux catégories d'organisations :

- 1a) les universités et d'autres établissements d'enseignement supérieur établis au titre du droit communautaire ou du droit d'un État membre;
- 1b) les organisations ou institutions de recherche scientifique établies au titre du droit communautaire ou du droit d'un État membre.

Dans le cas des « autres organisations », l'article 3 du Règlement n° 831/2002 prévoit l'approbation préalable du Comité du secret statistique, pour les organisations qui souhaitent soumettre des demandes d'accès à des données confidentielles à des fins scientifiques. La Décision 2004/452/CE de la Commission comporte une liste d'autres organisations qui ont été considérées comme admissibles. La condition préalable à l'admissibilité est que l'institution ait fait la démonstration qu'elle répond à un ensemble de critères. Le CSS a approuvé ces critères à sa réunion du 10 décembre 2004. Des services particuliers des institutions de l'UE qui mènent des activités statistiques, peuvent être considérés admissibles en tant que chercheurs, en vue d'accéder à des microfichiers confidentiels particuliers, à condition que des garanties équivalentes soient fournies. Cela fait suite au précédent établi avec la BCE et les Banques centrales d'Espagne et d'Italie. Des universités établies à l'extérieur de l'Europe peuvent aussi être considérées comme admissibles; la Cornell University (É.-U.) a été la première à être incluse dans cette liste. Les efforts se poursuivront en vue d'élargir la liste des autres organisations qui pourraient être considérées comme admissibles.

Établissement d'ententes bilatérales sur l'octroi de licences et la décentralisation des centres sécuritaires

Une composante importante de l'élaboration d'un nouveau système de protection de la confidentialité est l'établissement d'un réseau de centres sécuritaires. Pour le moment, le centre sécuritaire pour les ensembles de données mentionnées dans le Règlement n° 831/2002 de la Commission se trouve à Eurostat. Les responsables d'Eurostat discuteront avec les États membres de la possibilité de décentraliser le centre sécuritaire dans les États membres, par la voie d'ententes bilatérales, ou de créer les conditions nécessaires pour conclure des ententes de licence avec les institutions établies.

3. Enjeux méthodologiques

En général, les dispositions législatives au niveau des nations et de l'Europe sont relativement harmonisées en ce qui a trait aux données considérées comme confidentielles. Toutefois, lorsque ces dispositions législatives sont mises en œuvre, les critères utilisés diffèrent considérablement d'un pays à l'autre. Ils comportent parfois un poids historique important. Parfois aussi, ils n'ont pas de base scientifique solide, et dans nombre de cas, ils entraînent des solutions conservatrices parce que les risques réels ne sont pas bien maîtrisés.

La diversité des interprétations est une conséquence du fait qu'il n'existe pas d'approche harmonisée en ce qui a trait au risque de divulgation. Avant de tomber d'accord sur le risque de divulgation, il faut d'abord convenir de la nature délicate des données (dans quelle mesure les variables du fichier sont « privées »), ainsi que de la possibilité d'apparier ces données avec des données de sources externes, c'est-à-dire de la présence de variables clés ou de variables d'identification. En deuxième lieu, il faut déterminer une façon harmonisée de mesurer le risque. Des travaux méthodologiques sont nécessaires pour rapprocher les différents points de vue ou pour exprimer une préférence à l'égard de l'un d'eux.

Il est évident ici qu'on doit disposer de critères de base communs qui, tout en fournissant un niveau d'harmonisation satisfaisant, comportent le niveau de souplesse requis pour s'adapter à la perception particulière qu'a la société de chaque pays. Cela a aussi pour avantage que l'on dispose d'une base plus solide acceptée au niveau international, qui justifie dans une plus large mesure les choix effectués au niveau national quant à la diffusion des microdonnées.

Dans une certaine mesure, il en va de même lorsqu'Eurostat doit concevoir, de concert avec les États membres, des façons de rendre les microdonnées anonymes en vue de les diffuser aux chercheurs. Même s'ils partagent des objectifs communs :

- la nécessité de suivre les principes du règlement en ce qui a trait au droit à la protection de la vie privée;
- la nécessité de maintenir la confiance qu'ont les répondants à l'égard du système statistique;
- la nécessité de contrôler les diffusions, afin d'éviter les bris de confidentialité

les différences dans la perception du risque et l'absence d'une mesure universelle du risque rendent très faible la possibilité d'un consensus. Une part du problème vient de l'absence de connaissance du risque réel.

La situation serait améliorée si, encore une fois, les experts européens convenaient d'un ensemble de mesures et de seuils devant être utilisés par les praticiens. Cela nécessite probablement davantage de recherches comparatives concernant les mesures existantes et la mise au point des méthodes. En parallèle, d'autres recherches pourraient être effectuées quant à la mesure réelle du risque.

4. Perspectives pour l'avenir

En ce qui a trait aux perspectives à moyen terme, trois composantes principales sont déjà déterminées.

CDS dans le SSE

Établissement de centres de réseaux d'excellence pour le contrôle de la divulgation statistique. L'idée de ces centres découle du concept de partage plus efficace du travail entre les différentes institutions du SSE (Système statistique européen), grâce à des solutions organisationnelles et à un cadre institutionnel appropriés pour ces types modernes de coopération et de spécialisation du travail.

Dans ce dernier cas, le partage du travail entre les différents États membres créera des synergies, étant donné que chaque Institut National Statistique (INS) participant se concentrera sur des domaines particuliers, et que le produit de ce travail profitera à tous les INS, en dernier ressort, ce qui devrait entraîner l'augmentation de la qualité des statistiques du SSE. Il est en outre essentiel pour la production de données statistiques comparables d'un pays à l'autre et au niveau européen, que des méthodes et des outils similaires soient utilisés pour protéger la confidentialité des données publiées. Tant et aussi longtemps que les États membres compileront leurs statistiques au moyen de méthodes différentes de contrôle de la divulgation statistique (CDS), la compilation de statistiques européennes sera gravement compromise.

Le centre de réseaux pilote pour le CDS a été défini afin de tenir compte en premier lieu des objectifs suivants :

- établir des normes pour la protection des ensembles de microdonnées, selon les méthodes et les critères d'évaluation du risque de divulgation;
- améliorer les techniques de protection des données tabulées et élaborer des critères harmonisés;
- élargir et développer des outils logiciels de CDS, à la fois pour les microdonnées et les données tabulées, en vue de respecter les environnements particuliers de production et de diffusion du SSE.

Eurostat a l'intention d'évaluer et d'élaborer davantage l'approche des centres de réseaux, en vue d'harmoniser les pratiques de CDS dans le SSE, de promouvoir la définition et l'utilisation de pratiques éprouvées, de mettre à niveau les logiciels du CDS dans le SSE et d'assurer l'accès à distance aux microdonnées.

Fichiers à grande diffusion

Les fichiers à grande diffusion (FGD) sont les plus accessibles; il s'agit de produits de microdonnées largement et librement utilisés, qui sont mis à la disposition des institutions statistiques, mais leur valeur est limitée pour une part importante de la recherche pertinente au niveau stratégique. Néanmoins, ces fichiers sont utiles à des fins de recherche et comme outils d'enseignement, et ils constituent une bonne publicité pour un institut statistique. La distribution continue des fichiers à grande diffusion est compromise par le risque accru de réidentification lié aux percées technologiques au chapitre du logiciel de couplage et de la grande disponibilité des dossiers administratifs. Au cours de la dernière décennie, les chercheurs ont élaboré des méthodes de plus en plus poussées pour les produits de données restreints. L'élaboration d'une méthodologie pour la production de

données synthétiques ou virtuelles est une activité relativement récente. L'un des objectifs clés de la méthode consiste à préserver les représentations fidèles des données originales, afin que les inférences à partir des données synthétiques soient les plus uniformes possibles avec les inférences tirées des données originales. Parmi les autres caractéristiques attrayantes de l'approche des données synthétiques figure le fait qu'elle peut servir à créer des fichiers à grande diffusion multiples à partir des mêmes données sous-jacentes, en ciblant des publics différents. La méthodologie des fichiers synthétiques, comme mesure de remplacement des fichiers à grande diffusion, nécessite des recherches plus poussées.

On continuera de promouvoir les travaux au niveau sectoriel, en vue de la définition de critères pour la création de fichiers à grande diffusion (comme les travaux permanents du groupe de travail de l'EU-SILC sur l'anonymisation et la création de fichiers à grande diffusion), grâce à l'établissement de groupes de travail sectoriels qui définiront les fichiers à grande diffusion pour chaque enquête.

Accès à distance contrôlé aux microdonnées

Parmi les approches sensibles pour faciliter les recherches de grande qualité figure la conservation des données dans un environnement sécuritaire à accès à distance restreint.

L'accès à distance contrôlé a pour avantages qu'un chercheur n'a pas à se rendre dans un centre sécuritaire pour utiliser des données confidentielles et que les produits sont fournis relativement rapidement. Cette approche pour l'élaboration de procédures d'accès à distance, qui permet de réduire le fardeau pour les chercheurs, nécessite des investissements substantiels dans le matériel et le logiciel. Elle a fait une percée et est maintenant opérationnelle en Europe, aux Pays-Bas, au Danemark et en Suède. On étudiera avec les États membres la possibilité offerte par le septième programme de cadre de recherche dans le domaine des infrastructures de recherche, en plus de poursuivre l'élaboration d'une telle approche au niveau européen.

Certaines des exigences et certains des objectifs précisés dans les dispositions législatives ne sont pas fixes, mais évoluent avec le temps. Il est donc nécessaire que les INS et Eurostat passent en revue les pratiques et les méthodes de temps à autre. On a présenté certaines des pistes à suivre à plus long terme concernant la modification du cadre juridique actuel. En parallèle, on a décrit les axes concrets d'application qui rendent compte des orientations d'Eurostat à court et à moyen termes en ce qui a trait à la confidentialité. Eurostat souhaite créer des synergies utiles avec les experts et les INS, selon ces grandes lignes.

RÉFÉRENCES

Mercy, J.-L. et J. King (2003), "Developments at Eurostat for research access to confidential data" Joint ECE/Eurostat work session on statistical data confidentiality (Luxembourg, 7-9 April 2003) Working Paper 12.