



Conseil économique et social

Distr. générale
11 mars 2013
Français
Original: anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Soixante et unième réunion plénière

Genève, 10-12 juin 2013

Point 4 b) de l'ordre du jour provisoire

Comment les bureaux de statistique nationaux devraient-ils réagir, en passant du souci d'éviter les risques à la gestion des risques?

La gestion des risques en matière d'accès aux microdonnées: deux exemples canadiens

Note établie par Statistique Canada

Résumé

Le présent document donne un aperçu de la nature du processus adopté par le Canada pour commencer à permettre aux personnes considérées comme employés d'avoir accès aux microdonnées des entreprises tout en gérant les risques inhérents à ce processus, dans un petit pays comme le Canada. Ce processus a été adopté depuis l'inauguration du Centre canadien d'élaboration de données et de recherche économique. Le présent document portera sur les aspects juridiques qui ont motivé cette décision, les systèmes mis en place pour prévenir une utilisation inappropriée des données, les procédures adoptées pour garantir la sécurité des locaux, la nature du processus d'accréditation, la façon dont le contrôle de la divulgation est organisé et les axes envisagés pour l'évolution future. On y trouvera également des explications sur le processus d'accréditation des chercheurs qui souhaitent accéder aux données sur la société et les entreprises.

I. Introduction

1. L'organisme Statistique Canada est chargé d'analyser les données qu'il collecte mais le volume de ces données dépasse largement sa capacité d'analyse. Pour que les Canadiens tirent tout le bénéfice possible des informations qu'il a recueillies à leurs frais, Statistique Canada renforce ses capacités d'analyse en nouant des partenariats avec des universitaires et des organisations. Ses partenaires ont accès aux données grâce à divers moyens qui peuvent être vus comme un continuum couvrant les différents niveaux de risque. Des procédures appropriées de gestion des risques sont appliquées à chaque palier de ce continuum.

2. Ainsi, les fichiers de microdonnées à usage public font l'objet d'une analyse minutieuse avant d'être diffusés à grande échelle et les microdonnées confidentielles sont protégées par un contrôle direct étroit ou par un procédé technique comme l'accès à distance. En ce qui concerne l'accès direct aux microdonnées détaillées, Statistique Canada use des pouvoirs que lui confère la loi pour accréditer des chercheurs agréés en tant que personnes considérées comme ses employés qui procèdent à une analyse scientifiquement valide, conduite à des fins statistiques uniquement, appropriée pour un bureau national de statistique. En étant considérés comme des employés de Statistique Canada, ces chercheurs sont assujettis à la loi sur la statistique qui prévoit des peines lourdes en cas de non-respect de la confidentialité des informations données par les répondants.

3. L'accès de ces personnes aux microdonnées sur la société et les entreprises relève de politiques, de règles et de procédures qui le limitent aux chercheurs agréés des centres sécurisés, en particulier les centres de données de recherche (RDC) et le Centre canadien d'élaboration de données et de recherche économique (CDER). Bien que ces personnes considérées comme des employés encourent les mêmes sanctions juridiques que ses salariés, Statistique Canada a appliqué de nouvelles stratégies de gestion des risques de façon à préserver la confiance des répondants, qu'il s'agisse de particuliers ou d'entreprises.

4. Il arrive que les bureaux nationaux de statistique adoptent une stratégie d'accès minimal de façon à supprimer pratiquement tous les risques en matière d'accès aux microdonnées. Jusqu'à une époque récente, c'est cette méthode qui prévalait pour l'accès aux données sur les entreprises du Canada où, la population étant relativement peu nombreuse, il est assez aisé d'identifier les grandes entreprises. Toutefois, lorsqu'un processus d'élaboration des politiques repose sur un système qui soutient une recherche éthiquement responsable et sur une communauté de chercheurs professionnels comptant des spécialistes de l'analyse, chercher à éliminer tout risque n'est peut-être pas le meilleur moyen pour les bureaux nationaux de statistique d'aborder le problème. Le présent document décrit comment le Canada est passé du souci d'éviter les risques à la gestion des risques pour produire davantage d'analyses à partir des ensembles de données nationales et servir ainsi l'intérêt de tous les Canadiens. L'évolution récente de la situation canadienne, qui a conduit Statistique Canada au continuum, est illustrée par deux nouvelles activités: la simplification des procédures relatives aux chercheurs accrédités et l'accès aux microdonnées sur les entreprises.

II. Accréditation des chercheurs

5. L'accès aux microdonnées sociales des personnes considérées comme des employés, par le biais des RDC, suppose une série de propositions, d'examens, d'approbations et de procédures de sécurité ainsi que de documents à remplir avant que les chercheurs commencent à travailler sur leur projet. Pour déterminer les domaines où les risques sont moindres et réduire le poids des formalités dont doivent s'acquitter les chercheurs expérimentés de Statistique Canada ayant fait leurs preuves en matière de respect de la confidentialité, le Canada a élaboré une procédure d'accréditation à deux niveaux, le premier permettant de s'assurer que les chercheurs considérés comme des employés, qui ont accès aux microdonnées confidentielles du pays sont qualifiés et suffisamment expérimentés pour mener leurs travaux en toute indépendance et qu'ils suivent les procédures visant à préserver le caractère confidentiel des informations communiquées par les Canadiens; et le second, simplifiant la procédure de demande relative à un projet pour les chercheurs hautement qualifiés et très expérimentés en leur délivrant une accréditation valable dix ans et supprimant l'obligation de solliciter une accréditation pour chaque projet.

6. La procédure d'accréditation de niveau 1 existe depuis plus dix ans et s'applique à chaque soumission de proposition de recherche. Les informations fournies par le demandeur sont examinées et approuvées par des pairs – universitaires ou fonctionnaires, selon qu'il convient. Ce niveau d'accréditation donne accès aux données confidentielles de Statistique Canada, mais nécessite un examen complet par les pairs de chaque projet. Les demandeurs doivent remplir une demande dans laquelle ils justifient de leur expérience en matière de recherche et de leurs compétences techniques pour travailler sur les données demandées. De plus, l'étudiant du deuxième ou du troisième cycle est tenu de produire une lettre de soutien indiquant qu'un chercheur principal le conseillera ou l'assistera lorsque des questions d'ordre statistique ou analytique se poseront au cours de ses travaux. Les obligations propres à ce niveau d'accréditation sont énumérées au tableau 1.

Tableau 1

Conditions à remplir pour l'accréditation de niveau 1

Contributions les plus importantes à la recherche
Supervision d'étudiants de deuxième ou de troisième cycles
Appartenance à des organismes professionnels (par exemple, associations ou sociétés professionnelles, comités d'évaluation de subventions, etc.)
Projets de recherche financés
Chaires de recherche ou autres prix reconnaissant les réalisations
Présentations pertinentes à des groupes gouvernementaux ou non gouvernementaux
Discours sollicités
Publications dans des revues à comité de lecture, au cours des cinq dernières années, qui illustrent les compétences techniques

7. La procédure d'accréditation de niveau 2 est récente et vise à simplifier l'accès aux données pour les chercheurs expérimentés et compétents. Le chercheur doit avoir réalisé des travaux reconnus, être spécialisé dans un domaine ou une discipline et prouver qu'il possède les connaissances et l'expérience nécessaires pour traiter des informations personnelles confidentielles ainsi que les compétences techniques lui permettant de mener le genre de recherches conduites par les RDC de Statistique Canada. On trouvera, au tableau 2, la liste des conditions à remplir pour l'accréditation de niveau 2. Chacune des demandes sera examinée et approuvée par un groupe de collègues universitaires et de

cadres de Statistique Canada. Les chercheurs qui bénéficient d'une telle accréditation n'ont plus à soumettre leurs projets à cet examen.

Tableau 2

Conditions à remplir pour l'accréditation de niveau 2

Expérience dans le traitement d'informations confidentielles au sein d'un RDC; ou expérience équivalente; ou preuve selon laquelle le demandeur a suivi un programme en ligne sur la confidentialité et la sécurité
Indication de l'intérêt que présentent, pour la collectivité, les recherches actuelles ou précédentes du demandeur (par exemple, évolution des politiques ou des programmes publics, acquisition de nouvelles connaissances venant éclairer une question sociale ou économique d'actualité, etc.)
Contributions les plus importantes à la recherche
Supervision d'étudiants du deuxième ou du troisième cycle
Appartenance à des organismes professionnels (par exemple, associations professionnelles, comités d'évaluation des subventions, etc.)
Collaboration avec un institut ou un organisme de recherche
Projets de recherche financés
Chaires de recherche ou autres prix reconnaissant les réalisations
Présentations pertinentes à des groupes gouvernementaux ou non gouvernementaux
Discours sollicités
Analyses pour des revues spécialisées pratiquant l'examen collégial
Publications dans des revues à comité de lecture, au cours des cinq dernières années, qui illustrent les compétences techniques
Enseignement des méthodes de recherche

8. L'accréditation de niveau 1 a été introduite dans le programme des RDC dès le début, en 2000. Elle est désormais utilisée aussi par le nouveau Canadian Centre for Data Development and Economic Research (CDER) (voir ci-dessous). L'accréditation de niveau 2 a été approuvée; son application, qui relève d'une procédure en cours d'élaboration, devrait être effective en 2013/2014.

III. Centre canadien d'élaboration de données et de recherche économique

9. En octobre 2012, la création du Centre canadien d'élaboration de données et de recherche économique (CDER), au siège de Statistique Canada à Ottawa, a permis de donner un accès sécurisé aux microdonnées sur les entreprises à de nombreux chercheurs considérés comme employés qui travaillent sur des projets de recherche faisant l'objet d'un examen par des pairs. Auparavant, l'accès à ces microdonnées restait limité. Statistique Canada gérait un programme de bourses d'études pour de rares doctorants et postdoctorants. Des contrats étaient établis avec un nombre restreint d'universitaires dont les travaux profitaient directement aux programmes de Statistique Canada ou à leur élaboration. Quelques analystes d'autres organismes et services gouvernementaux, qui

travaillaient sur des projets prioritaires, y étaient accueillis. Compte tenu de ses moyens limités, Statistique Canada s'employait surtout à éviter les risques.

10. Contrôler la divulgation des données sur les entreprises est un défi, en particulier lorsque la demande porte sur un grand nombre de tableaux. La loi sur la statistique interdit de divulguer les détails des données fournies par les répondants. Compte tenu de l'asymétrie des données sur les entreprises et du faible nombre d'entreprises dans certains secteurs ou dans certaines régions, il est relativement facile de reconnaître des entités dans ces ensembles de données. En raison de ces deux caractéristiques (population restreinte et prépondérance d'un petit nombre d'observations dans les totaux), la qualité des données produites suppose un taux de réponse élevé et donc la solidité du partenariat instauré depuis longtemps par Statistique Canada avec ses répondants.

11. Ce partenariat ne repose pas seulement sur les clauses juridiques figurant dans la loi sur la statistique qui portent sur le droit du bureau à solliciter des données et à son obligation de les protéger. Il dépend aussi de la collaboration instituée de longue date entre le bureau et les répondants en vue de produire de précieuses informations statistiques. Lorsque de nouvelles enquêtes sont élaborées, le bureau explique aux répondants les raisons pour lesquelles elles doivent être menées et s'efforce de simplifier leur tâche. Ainsi, il reste en relation avec eux pour leur assurer que leurs données sont protégées en permanence.

12. Tout ceci est important si l'on veut préserver la qualité des données produites sur les entreprises. La perte de confiance de quelques entreprises répondantes – même peu nombreuses, en particulier s'il s'agissait des plus importantes – affecterait directement et de manière considérable les données sur les entreprises produites par le système statistique. L'ensemble des répondants pourrait alors ressentir une certaine ambivalence face à la communication de données et la qualité des informations produites par le système statistique s'en trouverait diminuée.

13. Le non-respect de la confidentialité des données sur les entreprises peut coûter très cher et ce risque est élevé car des personnes extérieures pourraient être tentées de décrypter les bases de données pour identifier les sujets concernés par les observations. Par appât du gain, les chercheurs peuvent faire mauvais usage des données sur les entreprises et ne pas en respecter le caractère confidentiel. C'est pourquoi, face au risque de non-respect de cette confidentialité et à l'importance des conséquences d'une divulgation des informations, Statistique Canada a depuis toujours opté pour une approche très ferme visant à minimiser les risques en matière de délivrance des autorisations d'accès aux microdonnées sur les entreprises.

14. Statistique Canada a toutefois reconnu la nécessité d'utiliser les microdonnées sur les entreprises à des fins d'analyse et les avantages qu'on pouvait manifestement en tirer. Il a donc développé, en interne, sa capacité de recherche économique au sein des divisions analytiques qui collaborent avec des chercheurs de la communauté fédérale et des milieux universitaires; ces partenariats sont toutefois limités pour ce qui est de l'accès fourni par les RDC aux données à caractère social. Statistique Canada a estimé que s'il autorisait les personnes considérées comme des employés à accéder aux données comme la loi l'y autorisait et s'il mettait en place des procédures de contrôle des risques liés à l'accès aux microdonnées, il pourrait tirer avantage de l'ouverture de l'accès aux microdonnées sur les entreprises. En faisant appel à des compétences et à des ressources extérieures – celles d'autres services gouvernementaux et d'instituts universitaires et de recherche – il serait mieux à même d'exécuter le mandat que lui confère la loi sur la statistique qui consiste à fournir des informations aux Canadiens en analysant «des renseignements statistiques sur les activités commerciales, industrielles, financières, sociales, économiques et générales de la population et sur l'état de celle-ci».

15. Le CDER a donc été créé pour permettre aux chercheurs hautement qualifiés considérés comme des employés de mener des recherches importantes sur les entreprises à l'appui de la politique et de son élaboration. Cette avancée devait néanmoins être mise en balance avec le risque accru de violation de la confidentialité induit par un plus large accès. C'est pourquoi, avant que la décision ne soit prise de créer le CDER, Statistique Canada a élaboré des politiques et des procédures de gestion du risque afin de maintenir celui-ci à un niveau acceptable. Les informations fournies ci-dessous portent essentiellement sur les mesures adoptées en plus de celles appliquées dans les RDC par le biais desquels Statistique Canada fournit un accès aux microdonnées sociales. Il s'agit de faire en sorte que les personnes considérées comme des employés travaillent dans des lieux sécurisés grâce à des procédures sûres, aux projets qui relèvent du mandat du bureau de statistique.

A. Nature du processus d'approbation

16. Comme c'est le cas pour les RDC, les projets soumis au CDER font l'objet de deux types d'examen: un examen par les pairs qui s'assure que les projets satisfont à certains critères de professionnalisme, et un examen institutionnel pour vérifier que les projets entrent bien dans le cadre du mandat de Statistique Canada et n'affectent ni sa capacité de continuer d'inspirer confiance aux répondants, ni sa réputation d'objectivité ou de neutralité. Les chercheurs associés aux projets se soumettent également à une procédure d'accréditation en vue d'évaluer leur capacité de mener lesdits projets. Tout projet doit faire l'objet de ces deux examens et les chercheurs qui y sont associés doivent être accrédités pour le projet en attente d'approbation.

17. L'examen par les pairs et la procédure d'accréditation peuvent être menés de deux façons mais toujours par des chercheurs des milieux universitaires, ce qui permet à Statistique Canada de faire appel à des compétences extérieures pour l'aider à accorder l'accréditation. Lorsque le projet a été examiné par un organisme officiel de financement de la recherche – comme le Conseil de recherches en sciences humaines – et que celui-ci lui a octroyé des fonds, Statistique Canada en déduit que le projet et les chercheurs qui lui sont associés satisfont aux critères professionnels requis. Lorsque le projet n'a pas reçu de fonds d'un tel organisme son intérêt professionnel et les qualifications des chercheurs sont évalués par deux membres d'un comité d'experts extérieurs composé essentiellement d'universitaires nommés par Statistique Canada.

18. C'est au stade de l'examen institutionnel que la procédure diffère entre le CDER et les RDC. L'examen institutionnel du projet est réalisé par un comité constitué de cadres expérimentés du bureau et de spécialistes des questions traitées qui examinent les propositions des RDC. Il sert à vérifier que: les recherches ne serviront qu'aux fins indiquées dans la proposition de projet; les chercheurs ont mentionné toutes leurs sources de financement et de soutien; nulle autre personne ou organisation n'a d'«intérêt» dans le projet. À cette fin, il est demandé aux chercheurs de fournir des renseignements sur:

- a) L'origine du soutien financier ou de l'aide en nature qu'ils reçoivent pour réaliser leur projet;
- b) Les travaux qu'ils ont déjà réalisés ou qu'ils entendent mener sur la population d'entreprises qu'ils souhaitent étudier dans leur projet actuel;
- c) L'accès qui leur est donné à d'autres microdonnées sur les entreprises par Statistique Canada ou dans d'autres pays;
- d) Les contacts qu'ils entretiennent avec d'autres chercheurs travaillant sur des microdonnées similaires ou connexes à Statistique Canada;

e) Leurs liens avec des organismes dont dépendent administrativement les entreprises sur lesquelles portent les données auxquelles ils auront accès; et

f) Leurs liens avec les entités qui pourraient tirer un avantage financier des informations relatives aux entreprises auxquelles les chercheurs ont accès.

19. Comme il a été évoqué précédemment, les avantages financiers qu'un chercheur est susceptible de tirer d'une utilisation inappropriée des données sur les entreprises représentent un risque qui doit être géré par un examen attentif des contacts des chercheurs qui utilisent les microdonnées. Une fois le projet approuvé et le chercheur accrédité, ce dernier fait l'objet de la procédure visant à l'instituer personne considérée comme un employé au sens de la loi sur la statistique.

B. Systèmes mis en place pour empêcher l'identification des données des entreprises

20. Si la loi sur la statistique interdit de communiquer les données obtenues des répondants, le contrat implicite qui lie le bureau avec ses répondants dépend de nombreux accords portant sur les informations qu'ils fournissent. Les renseignements transmis par une entreprise pourraient être utilisés à son encontre par d'autres biais que la publication involontaire par Statistique Canada de chiffres individualisés. La possibilité offerte aux chercheurs considérés comme employés qui sont accrédités pour un projet précis d'accéder aux bases de microdonnées sur les entreprises et qui sont susceptibles d'intégrer ces informations dans leurs travaux personnels constitue l'un des risques qui a fait l'objet de procédures opérationnelles.

21. Comme dans le cas des microdonnées sociales et des RDC, le caractère confidentiel de toutes les données produites par le CDER a été vérifié avant qu'on soit autorisé à les sortir des locaux. En dépit de cette vérification obligatoire et du fait que les chercheurs considérés comme employés n'ont pas le droit de les imprimer, il est possible que des informations confidentielles soient sorties du CDER par divers moyens, qui ne pourraient être contrôlés que par des mesures radicales inacceptables. En outre, on ne peut empêcher le prélèvement de données sur les blocs de mémoire de certaines personnes.

22. Pour remédier à ce problème, un système d'accès en deux temps a été élaboré pour les données sur les entreprises. Dans un premier temps, les chercheurs se voient autorisés à utiliser une base de données remaniée. L'algorithme utilisé pour produire cette base de données croise les données concernant l'une ou l'autre entreprise de façon aléatoire (l'entreprise A pourrait se voir attribuer les recettes de l'entreprise B et les actifs de l'entreprise C), tout en conservant dans une certaine mesure les corrélations et les autocorrélations entre les variables continues, certaines équations comptables et la distribution marginale de chaque variable. À partir de cette base remaniée, les chercheurs peuvent étudier la structure de la base de données et concevoir leurs propres programmes informatiques.

23. Les chercheurs considérés comme employés sont ensuite autorisés à utiliser les données réelles mais sans possibilité d'interaction, c'est-à-dire qu'ils soumettent leurs programmes à un mode de traitement par lots. Avec ce procédé, ils ne peuvent pas visualiser les données réelles et l'utilisation qu'ils font des données auxquelles ils ont accès peut être suivie et enregistrée d'une façon plus facile à contrôler. De plus, les chercheurs acceptent de ne pas essayer de visualiser les observations individualisées. Ils sont également informés que ce qu'ils tapent sur le clavier est enregistré et que des contrôles aléatoires seront réalisés pour garantir qu'ils ne se livrent à aucune activité inappropriée. Enfin, ils acceptent d'informer immédiatement le CDER s'il leur arrive, par inadvertance, d'accéder à des données individuelles.

C. Institution signataire des contrats de recherche

24. Qu'ils travaillent avec les RDC ou le CDER, les chercheurs prêtent serment avant d'entrer en fonction, comme le prescrit la loi sur la statistique au point 6 1), en vertu de laquelle ils sont soumis aux mêmes obligations et passibles des mêmes peines que les autres employés de Statistique Canada. Ils acceptent également de respecter toutes les politiques de sécurité énoncées dans le manuel de sécurité du bureau. Outre les peines prévues par la loi, les employés ordinaires de Statistique Canada risquent, en cas de non-respect de la confidentialité des données, de perdre leur emploi sans possibilité d'être réengagés dans le service public. Les chercheurs, eux, risquent d'être exclus à vie des RDC et du CDER.

25. Pour donner plus de poids aux peines et aux sanctions prévues en cas de violation de la loi sur la statistique et d'autres politiques de Statistique Canada, les chercheurs signent un contrat de recherche et, en ce qui concerne les chercheurs du CDER, ce contrat est contresigné par leur employeur. Si l'employeur n'est pas tenu de sanctionner le chercheur qui ne respecte pas les termes de son contrat, il est toutefois informé des obligations auxquelles son employé doit se soumettre pour accéder aux données et de la gravité des faits en cas de non-respect de celles-ci. Il a dès lors l'obligation morale de réagir au comportement aberrant de son employé au cas où celui-ci enfreindrait la loi. Ainsi le contrat constitue-t-il un moyen supplémentaire de dissuader les chercheurs considérés comme employés de se lancer dans des activités interdites alors qu'ils sont employés par des institutions qui apprécient la possibilité qui leur est offerte de détacher des chercheurs auprès de Statistique Canada pour y utiliser des microdonnées.

D. Orientations futures de l'accès aux données sur les entreprises

26. La gestion des risques a été l'un des facteurs déterminants des choix du CDER (ses procédures et son implantation au siège) concernant le continuum des modes d'accès aux données, mais il n'a pas été le seul. Ont aussi été pris en considération l'état de préparation des données sur les entreprises à des fins de recherche, le coût et la complexité de la mise en place d'un système sécurisé qui permettrait d'accéder aux données à partir de différents lieux du pays et l'idée que Statistique Canada devait faire ses preuves avec un accès limité avant offrir un accès plus large. Tout ceci a conduit Statistique Canada à implanter le CDER dans les locaux de son siège à Ottawa. Comme il fallait que les experts participent à la mise en place et à la maintenance des bases de données, le CDER a été installé à proximité de la principale unité de recherche et d'analyse économiques du bureau, la Division de l'analyse économique, dont il constitue un élément.

27. Comme la documentation des microdonnées sur les entreprises atteint désormais les normes fixées pour celle des données sociales et que l'on envisage de plus en plus des partenariats avec des universitaires pour couvrir le coût de l'expansion d'un accès sécurisé, on s'attend à ce que l'emplacement du CDER sur le continuum d'accès soit modifié. Si le CDER peut prouver aux Canadiens que les politiques, les procédures et les systèmes mis en place sont efficaces pour gérer les risques et maintenir la confiance des milieux d'affaires, il se pourrait que l'accès aux données sur les entreprises soit relevé au sein du continuum. Un moyen essentiel pour susciter la confiance sera de montrer que la recherche produite par le CDER peut fournir des informations importantes aux Canadiens et qu'elle sert les intérêts des répondants.

E. Conclusions

28. Chez Statistique Canada, des procédures de gestion des risques ont été appliquées à différents niveaux du continuum d'accès. Il est clair qu'il faut mettre en place des stratégies de gestion des risques pour maintenir la confiance des répondants. Une mise en œuvre attentive des politiques, règles et procédures peut aider les bureaux nationaux de statistique à gérer les risques liés à l'accès aux données et permettre à un accès progressivement plus ouvert dans l'intérêt de tous. Ces deux exemples récents concernant le Canada montrent qu'il est possible pour les bureaux nationaux de statistique de progresser dans le continuum pour passer du souci d'éviter les risques à la gestion des risques, afin de faciliter l'accès aux données des analystes professionnels qui intègrent Statistique Canada en tant que chercheurs considérés comme employés.
