



**Conseil économique  
et social**

Distr.  
GÉNÉRALE

CES/2003/12  
7 avril 2003

FRANÇAIS  
Original: ANGLAIS

---

COMMISSION DE STATISTIQUE et  
COMMISSION ÉCONOMIQUE POUR L'EUROPE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

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

**ACCÈS AUX MICRODONNÉES – QUESTIONS À CONSIDÉRER,  
ORGANISATION ET DÉMARCHES**

Document présenté par l'Australian Bureau of Statistics (ABS)<sup>1</sup>

**I. RAPPEL DES FAITS**

1. S'il importe d'assurer la confidentialité des données, c'est non seulement pour des raisons juridiques et éthiques mais aussi pour conserver la confiance du public. Or, la demande croissante de données détaillées, conjuguée à la puissance et la capacité de plus en plus grandes des moyens techniques, ainsi que la disponibilité d'ensembles de données pouvant être mis en correspondance, rendent plus ardue la protection de la confidentialité des microdonnées. Il est pratiquement impossible de publier des microdonnées contenant des informations utiles sans courir aucun risque d'identification. Les séries de données longitudinales accentuent encore le problème. Nous ne pouvons plus nous fier uniquement aux différentes formes de modification des données pour maîtriser les risques de divulgation.

2. Par ailleurs, on observe une demande croissante de microdonnées pour étayer toute une gamme d'études et d'analyses de données secondaires. Grâce au développement de la puissance des ordinateurs, les chercheurs sont mieux à même d'entreprendre ce genre d'analyses.

---

<sup>1</sup> Établi par Dennis Trewin.

3. Plusieurs considérations amènent à s'interroger sur la meilleure manière de rendre les microdonnées accessibles aux chercheurs:

- Des données de grande valeur (et de bonne qualité) sont sous-utilisées;
- Les chercheurs tentent parfois de se procurer des microdonnées par la collecte d'ensembles de données de substitution, ce qui représente un gaspillage des ressources publiques (pour obtenir en fin de compte des données de qualité probablement médiocre);
- Il arrive que les organismes publics fassent appel à des sources de données de remplacement pour obtenir des données d'enquête à des fins d'études et d'analyses, les données ainsi collectées étant de qualité inférieure et parfois inutilisables pour l'établissement des statistiques nationales.

4. Nous devons également prendre en compte un autre élément important – le potentiel analytique incroyable que présentent les ensembles de données raccordés entre eux, y compris ceux qui sont mis en correspondance avec les ensembles de données de l'ABS.

5. Ces divers facteurs nous ont amenés à repenser la manière dont nous assurons l'accès aux microdonnées. Cette réflexion peut également intéresser de nombreux autres services nationaux de statistique, dont beaucoup sont en train de modifier leurs méthodes de travail. Les mesures que nous avons prises, ou que nous envisageons de prendre, sont exposées dans le présent document. Il peut s'avérer nécessaire de différencier les stratégies mises en œuvre dans le cas des enquêtes auprès des ménages, d'une part, et des enquêtes auprès des entreprises, d'autre part. Dans le présent document, nous avons simplement tenté de décrire la situation de l'ABS mais en espérant que cet exposé pourrait s'appliquer à la situation de bien d'autres services nationaux de statistique.

6. Avant de poursuivre, il convient de souligner que quelles que soient les initiatives adoptées, elles doivent être à la fois légales et acceptables aux yeux du public. La loi peut certes être modifiée mais ce n'est pas la solution la plus rapide ou la plus directe et elle pourrait être la source de préoccupations inutiles. C'est pourquoi notre démarche s'inscrit dans le cadre de la législation en vigueur.

## **II. BRÈVE DESCRIPTION DE LA DEMANDE**

7. Dans des conditions idéales, les desiderata des utilisateurs seraient les suivants:

- Pouvoir exploiter les données de manière interactive;
- Avoir la possibilité de consulter les experts de l'ABS et disposer d'une bonne documentation descriptive des données;
- Avoir à leur disposition un nombre croissant d'ensembles de données;

- Se procurer des données de bonne qualité pour les populations et les variables qui les intéressent et quelques informations sur les sources d'erreurs;
- Obtenir des données publiées en temps voulu; et
- De plus en plus, avoir accès à des ensembles de données raccordés entre eux (y compris des données chronologiques liées).

8. Les chercheurs nous ont fait savoir à plusieurs reprises qu'ils jugeaient notre approche de la publication des microdonnées trop conservatrice. Pour répondre à leur demande, nous nous sommes concentrés récemment sur la manière d'élargir l'accès aux microdonnées, tout en préservant notre solide réputation en matière de protection des données personnelles (et de respect de la loi) qui est d'une importance capitale pour continuer à bénéficier d'un degré élevé de participation à nos enquêtes.

9. L'utilisation des ensembles de données raccordés conduit à envisager la possibilité de confier à l'ABS la garde d'autres ensembles de données que les siens pour s'assurer que leur confidentialité est protégée comme il se doit. Si une telle solution est entièrement compatible avec les objectifs d'un service national de statistique, il n'en demeure pas moins nécessaire d'élaborer des politiques et modalités pratiques appropriées à cette fin. Voir à ce sujet les paragraphes 24 à 28 du présent document.

### **III. MOYENS DE SATISFAIRE LA DEMANDE**

10. Il existe toute une gamme d'options, c'est-à-dire de moyens de diffusion, qui offrent un degré variable de protection contre les risques de divulgation de données confidentielles. Les premières options énumérées ci-après s'appuient en général davantage sur des données sécurisées, tandis que les dernières reposent plutôt sur un environnement sécurisé, notamment le recours à des engagements juridiquement contraignants qui prévoient des sanctions graves en cas d'infractions.

11. L'accessibilité des données et leur facilité d'emploi pour les chercheurs varient aussi selon l'option retenue.

12. La diffusion des microdonnées, qui est le thème précis du présent débat, est un élément clef à considérer lorsque l'on traite de l'accès aux données à des fins de recherche. La législation en matière de statistique nous autorise à diffuser des microdonnées mais uniquement «d'une manière qui ne soit pas susceptible de permettre l'identification de la personne ou de l'entité à laquelle ces données se rapportent». Des engagements sont également exigés. Néanmoins, il y a plusieurs moyens d'assurer l'accès aux microdonnées tout en respectant cette prescription légale. On trouvera ci-après des explications à ce sujet, de même qu'un résumé des options possibles dans le tableau 1.

13. Un groupe d'experts chargé de l'étude des microdonnées a été constitué pour nous aider à déterminer si le risque de divulgation est faible au point d'être acceptable (c'est-à-dire qu'il «ne soit pas susceptible de permettre l'identification») pour les modes de diffusion dans lesquels des microdonnées entrent en jeu. Ce groupe d'experts a examiné les deux principales catégories de risque ci-après:

- prévention de l'identification spontanée; et
- prévention du risque de mise en correspondance de données.

14. De l'avis des experts en droit, un engagement juridiquement contraignant de s'abstenir de certaines actions répond au critère selon lequel la diffusion de microdonnées doit se faire «d'une manière qui ne soit pas susceptible de permettre l'identification» et devrait entrer en ligne de compte quand on évalue le risque de divulgation.

15. Les avantages et inconvénients de chaque mode de diffusion seront étudiés plus en détail dans les sections suivantes.

**Tableau 1: Mode de diffusion de données destinées à étayer les recherches**

Mode de diffusion	Notes explicatives
1. Produits statistiques standard	Généralement présentés sous la forme de tableaux. Ce mode de diffusion restreint la nature des analyses que les utilisateurs peuvent entreprendre.
2. Cubes de données	Ce mode de diffusion fournit plus de détails et laisse aux chercheurs toute latitude pour élaborer leurs propres tableaux.
3. Services spéciaux de communication de données	Fournis à la demande des chercheurs, généralement moyennant un coût marginal.
4. Fichiers d'enregistrements unitaires confidentialisés (CURF)	Les données ne sont pas identifiables. Elles sont communiquées sur CD-ROM. Correspond à ce que l'on entend généralement par «publication de microdonnées».
5. Laboratoire de données avec accès à distance (RADL)	Accès à des fichiers d'enregistrements unitaires confidentialisés mais en ayant la possibilité d'obtenir des données plus détaillées grâce au contrôle plus strict qui peut s'exercer pour empêcher la mise en correspondance des informations communiquées avec celles de bases de données extérieures.
6. Laboratoire de données sur le site de l'ABS	Ne permet encore d'avoir accès qu'à des données non identifiables.
7. Collaboration	L'accès aux microdonnées se fait par l'intermédiaire d'un employé de l'ABS et non pas directement.
8. Analyse interne	En fait, cette formule confère au chercheur le statut d'un employé de l'ABS travaillant dans les locaux de cet organisme. Elle n'est praticable que si le chercheur aide effectivement l'ABS dans l'exécution de ses tâches.

## 16. Produits statistiques standard

- De quoi s'agit-il? – Ce sont des produits statistiques publiés, généralement sous forme de tableaux, sur papier et/ou sur support électronique.
- Protection de la confidentialité – Il s'agit de données sécurisées, auxquelles s'appliquent les pratiques normalisées de l'ABS en matière de confidentialité.
- Avantages – Commodité et prix peu élevé. Donnent une bonne indication de toute la gamme de données que l'on peut obtenir auprès de l'ABS. La disponibilité croissante de données électroniques téléchargeables améliore la facilité d'emploi de ces données aux fins d'une analyse ultérieure. Ce mode de diffusion est aisément accessible à tout un éventail de chercheurs. Il ne coûte pas cher à l'ABS.
- Inconvénients – La nature des analyses qui peuvent être entreprises sur cette base est limitée. Ces produits ne permettent pas de procéder à une analyse se fondant sur des microdonnées.
- Situation actuelle – On s'attache à élargir l'offre de données disponibles sous cette forme sur le site Web et à améliorer les métadonnées disponibles à titre complémentaire.
- Quand convient-il d'utiliser ce mode de diffusion? – Il ne faudrait pas sous-estimer l'intérêt des produits considérés en tant qu'outils de recherche commodes. Ce mode de diffusion devrait demeurer un facteur clef de la stratégie de diffusion pour tous les produits statistiques.

## 17. Cubes de données

- De quoi s'agit-il? – On entend par «cubes de données» la publication de matrices statistiques détaillées préalablement confidentialisées. Ce mode de diffusion offre une bonne solution lorsque la protection de la confidentialité peut être automatisée, en particulier pour les petites cases de tableaux (par exemple, dans le cas des recensements de la population). Les dispositions particulières en matière de confidentialité s'appliquant aux données sur le commerce permettent aussi de publier les données en question sous cette forme.
- Protection de la confidentialité – Il s'agit de données sécurisées, auxquelles s'appliquent les pratiques normalisées de l'ABS en matière de confidentialité (sauf pour les catégories de données soumises à des règles de confidentialité particulières, comme les données sur le commerce).
- Avantages – Accès relativement facile à des données plus détaillées que celles contenues dans les produits statistiques standard.
- Inconvénients – Les mêmes que pour les produits statistiques standard. En outre, la conception de cubes de données de bonne qualité n'est pas simple. Certains chercheurs éprouvent également des difficultés à utiliser les cubes de données. On ne pourra pas produire des cubes de données confidentialisés pour beaucoup de produits statistiques.

- Situation actuelle – Nous élargissons lentement la gamme de cubes de données disponibles. Nous nous attachons en outre à améliorer le savoir-faire des concepteurs de cubes de données.
  - Quand convient-il d'utiliser ce mode de diffusion? – Il se prête généralement mieux à la diffusion des données sur les personnes que sur les entreprises. En ce qui concerne certains produits statistiques, il faudrait l'intégrer dans la stratégie de diffusion.
18. Services spéciaux de communication de données
- De quoi s'agit-il? – Ce sont des produits statistiques diffusés – pas nécessairement sous la forme de tableaux – à la demande des chercheurs.
  - Protection de la confidentialité – Ces données sont sécurisées – Elles ne sont pas communiquées aux chercheurs si leur confidentialité n'est pas préalablement assurée.
  - Avantages – Les données et la forme sous laquelle elles sont présentées peuvent être adaptées aux besoins des utilisateurs.
  - Inconvénients – Ce mode de diffusion est coûteux pour certains chercheurs (de même que pour l'ABS). L'analyse est limitée par l'impossibilité de travailler de manière interactive. Le chercheur ne peut pas effectuer ses propres ajustements des microdonnées (par exemple, dans le cas d'observations aberrantes). L'exécution des différentes phases d'analyse des données peut s'avérer assez lente.
  - Situation actuelle – Cette formule est proposée à titre de service mais la demande n'est pas très forte. Nous ne cherchons pas à développer l'offre de ce genre de services sauf à l'intention de nos principaux clients et dans des domaines précis (par exemple celui des statistiques régionales).
  - Quand convient-il d'utiliser ce mode de diffusion? – Habituellement pour les statistiques présentées en tableau lorsqu'elles ne sont pas fournies sous la forme de produits standard et que l'accès aux microdonnées n'est pas possible. D'autres formes d'analyse s'effectuent plus généralement dans le cadre d'une formule de collaboration (voir ci-après).
19. Fichiers d'enregistrements unitaires confidentialisés (CURF)
- De quoi s'agit-il? – Ce mode de diffusion consiste à communiquer sur CD-ROM des fichiers de microdonnées qui ont été modifiés de manière à rendre improbable l'identification d'un individu ou d'une entité.
  - Protection de la confidentialité – Le Groupe d'experts chargé de l'étude des microdonnées donne des conseils quant aux ajustements à opérer pour protéger la confidentialité des données, ce qui peut nécessiter la mise en œuvre de méthodes de modification des données telles que la suppression de certaines variables, la limitation du degré de détail de certaines variables (en particulier d'ordre géographique), l'élimination, dans les données, de certains individus aisément identifiables et l'introduction de perturbations aléatoires. La confidentialité est également protégée en exigeant, de la part de tous les chercheurs qui

ont accès aux microdonnées, un engagement juridique aux termes duquel ils sont tenus de préserver le secret de ces données. En cas d'infractions, des sanctions (y compris le refus de lui fournir ultérieurement des microdonnées) seront prises à l'encontre du chercheur et éventuellement de l'institution dont il relève. Une procédure judiciaire est également envisageable.

- Avantages – Grande souplesse et commodité pour le chercheur.
  - Inconvénients – Le degré de détail recherché n'est pas toujours disponible. En général, on ne dispose pas de CURF pour les données concernant les entreprises. On a observé un petit nombre d'infractions aux engagements qui avaient été contractés (mais pas d'identification de données individuelles).
  - Situation actuelle – Ce mode de diffusion conservera son importance en tant qu'instrument dans le cadre de travail de la recherche et de l'analyse de données secondaire. La demande est forte. Nous nous efforçons d'améliorer l'actualité des données que nous communiquons.
  - Quand convient-il d'utiliser ce mode de diffusion? – Les statistiques ainsi diffusées sont considérées comme l'un des produits standard dans le cas des enquêtes auprès des ménages. On utilise sélectivement cette formule de diffusion pour d'autres données d'enquête qui conservent leur utilité à des fins de recherche même après que les méthodes de protection de la confidentialité leur ont été appliquées.
20. Laboratoire de données avec accès à distance (RADL)
- De quoi s'agit-il? – Des analyses demandées par des utilisateurs autorisés, par le biais de l'Internet, sont exécutées à l'aide des données contenues dans les fichiers d'enregistrements unitaires confidentialisés de l'ABS, et les résultats des analyses sont envoyés après avoir procédé à des contrôles largement automatisés de la confidentialité. Il s'agit d'un mode de diffusion semblable aux CURF si ce n'est qu'il devrait être possible d'offrir un accès à des données plus détaillées, le risque de mise en correspondance étant mieux contrôlable dans la mesure où les données ne quittent pas l'enceinte de l'ABS. Il est limité à la gamme de logiciels analytiques disponibles dans le contexte du Laboratoire de données avec accès à distance (SAS, SPSS, par exemple). Les produits seront inspectés manuellement avant d'être transmis aux utilisateurs.
  - Protection de la confidentialité – S'effectue selon l'avis du Groupe d'experts chargé de l'étude des microdonnées. L'inspection manuelle des produits est étayée par un dispositif de déclenchement automatique permettant d'identifier les produits qui peuvent exiger une inspection rigoureuse. L'ABS conserve les pistes de vérification et les dossiers des contrôles effectués. Des engagements juridiques devront être pris et les contrevenants seront passibles de sanctions.
  - Avantages – Accès à des données plus détaillées, ainsi qu'à des logiciels d'analyse que les chercheurs pourraient ne pas avoir à leur disposition. Services gratuits de traitement des données.

- Inconvénients – Manque de commodité par comparaison avec les CURF. Quelques retards dans la communication des produits. Gestion plus coûteuse pour l'ABS.
- Situation actuelle – Ce mode de diffusion a été lancé en avril 2003. Il sera modifié en fonction des réactions des utilisateurs. Le nombre d'ensemble de données disponibles par le biais de ce mécanisme va augmenter au fil du temps.
- Quand convient-il d'utiliser ce mode de diffusion? – Il sera utilisé à la place des CURF dans les cas où le risque de mise en correspondance des données que présentent ces derniers est trop élevé, et lorsqu'il s'avère trop risqué de s'appuyer uniquement sur des engagements/sanctions. Il pourra être utilisé, par exemple, pour les fichiers de données raccordés, notamment si l'un de ces fichiers est disponible à l'extérieur de l'ABS.

#### 21. Laboratoire de données sur le site de l'ABS

- De quoi s'agit-il? – Ce système est semblable au laboratoire de données avec accès à distance (RADL) si ce n'est qu'il n'est pas possible de télécharger des données d'enregistrements unitaires (ce qui peut se faire dans le système RADL pour un maximum de 30 enregistrements dans le but de détecter les valeurs aberrantes, etc.). Il convient de noter que notre situation est différente de celle de nombreux autres pays où il suffit de s'engager à garder le secret pour pouvoir avoir accès sur place à des fichiers d'enregistrements unitaires dont la confidentialité n'a pas été assurée. Dans le cadre de l'ABS, cela n'est autorisé que si le chercheur assiste effectivement le statisticien dans l'exercice de ses fonctions et si sa situation au regard de l'emploi permet de le considérer comme un employé de l'ABS, ce qui implique le paiement d'une rémunération.
- Protection de la confidentialité – Système semblable au RADL si ce n'est qu'un contrôle plus strict est exercé sur les produits; aucune donnée d'un enregistrement unitaire ne peut quitter les locaux de l'ABS.
- Avantages – Accès à des données qui ne pourraient sans doute pas être obtenues par le biais des CURF ou du RADL (par exemple des fichiers de données longitudinales). Accès plus direct aux experts de l'ABS.
- Inconvénients – Obligation de travailler dans les locaux de l'ABS. Gestion coûteuse pour le personnel de l'ABS, en particulier sur l'ensemble des neuf bureaux de cet organisme.
- Situation actuelle – Ce système est effectivement disponible à l'heure actuelle. Il a essentiellement été utilisé pour les fichiers de données longitudinales, notamment lorsque l'unité d'échantillonnage ou une partie des données provient du système administratif d'un autre organisme.
- Quand convient-il d'utiliser ce mode de diffusion? – Uniquement lorsque les services fournis par les CURF ou le RADL sont jugés inadaptés à un ensemble de données particulier ou lorsque le chercheur préfère cette formule et que l'ABS accepte de la mettre en œuvre dans le cas considéré.



## 22. Collaboration

- De quoi s'agit-il? – Cette formule consiste à travailler en collaboration avec un chercheur pour élaborer un produit (généralement une publication statistique) qui présente un intérêt pour l'ABS. Le produit statistique en question peut être (ou ne pas être) publié par l'ABS. Quoi qu'il en soit, cela n'empêche généralement pas les chercheurs de publier ou de présenter les résultats de ce travail ailleurs, notamment dans des revues scientifiques.
- Protection de la confidentialité – Le chercheur collaborant avec le personnel de l'ABS n'a pas directement accès aux données des enregistrements unitaires mais il doit passer par l'intermédiaire d'un employé de l'ABS avec lequel il travaille.
- Avantages – Profit mutuel découlant de la collaboration. Réel transfert de connaissances. Le chercheur a la possibilité de travailler la plupart du temps en dehors des locaux de l'ABS. Il arrive que des ressources financières soient mises à la disposition du chercheur pour ses travaux. Les dépenses encourues par le chercheur sont généralement plus faibles. Cette formule permet en principe d'avoir accès aux subventions accordées par le Conseil australien de la recherche.
- Inconvénients – Pas d'accès direct aux données. Ne vaut que pour les projets de collaboration qui présentent un intérêt pour l'ABS.
- Situation actuelle – La politique concernant les modalités de collaboration a été mise en place. Un Service des analyses, dont les effectifs se chiffrent à une trentaine de personnes, a été créé et est entré en activité il y a quatre ans, ce qui a véritablement permis d'organiser la collaboration avec la communauté des chercheurs qui, auparavant, s'effectuaient au cas par cas.
- Quand convient-il d'utiliser ce mode de diffusion? – Dans les cas où les résultats d'une collaboration s'avèrent avantageux tant pour les chercheurs que pour l'ABS. Pour certains projets revêtant un degré plus élevé de priorité, l'ABS recherche parfois le concours de collaborateurs. En dehors des règles de confidentialité, les principes qui devraient régir les activités de collaboration sont la conformité aux normes applicables aux marchés publics, la valeur statistique des projets, l'impartialité et la transparence, ainsi que la protection de la réputation de l'ABS.

## 23. Analyse interne

- De quoi s'agit-il? – L'ABS peut embaucher des personnes en qualité d'«agents» à condition qu'elles remplissent des fonctions utiles pour l'ABS dans ses activités. Les agents en question peuvent alors avoir accès aux données des enregistrements unitaires de l'ABS à condition de respecter les mêmes dispositions en matière de secret des données que les autres employés de l'ABS. Cette formule peut convenir lorsque l'ABS souhaite élaborer un produit pour lequel il lui manque certaines compétences et que le chercheur peut combler cette lacune. De manière générale, des dispositions peuvent être prises pour permettre aux chercheurs de publier certains aspects de leurs travaux en dehors de l'ABS avec l'autorisation de ce dernier.

- Protection de la confidentialité – Les agents susmentionnés sont soumis aux mêmes règles en matière de secret que les autres employés de l'ABS. Ils sont passibles de graves sanctions en cas d'infraction à ces règles.
- Avantages – Cette formule permet aux chercheurs d'avoir accès aux données des enregistrements unitaires. Elle est mutuellement avantageuse.
- Inconvénients – Une large part des travaux doit s'effectuer dans les locaux de l'ABS. La collaboration est limitée aux sujets qui intéressent directement l'ABS. Certaines restrictions s'appliquent aux résultats de la recherche. Il n'est pas toujours possible d'employer un chercheur en tant qu'agent de l'ABS.
- Situation actuelle – Cette formule a rarement été utilisée. Aujourd'hui, il est plus facile de la mettre en œuvre grâce aux modifications récentes apportées aux contrats de services publics.
- Quand convient-il d'utiliser ce mode de diffusion? – Quand l'ABS prend l'initiative d'engager un chercheur pour l'assister dans ses activités statistiques. (Bien évidemment, cette formule peut toujours offrir des avantages mutuels.)

#### **IV. ENSEMBLES DE DONNÉES RACCORDÉS**

24. Les ensembles de données raccordés constituent un exemple particulier de microdonnées auxquelles les utilisateurs peuvent souhaiter avoir accès. Je me réfère en l'occurrence à l'utilisation de méthodes de mise en correspondance des données qui permettent de combiner des enregistrements unitaires pour obtenir un ensemble d'enregistrements composite. Ce dernier peut être basé sur un appariement à l'aide d'identificateurs (ce que l'on appelle un «hard match») ou sur un appariement statistique faisant appel à une combinaison de variables (caractéristiques géographiques, âge, sexe, caractéristiques du ménage, etc.). Ces deux méthodes sont la source de difficultés sur le plan de la confidentialité. Les appariements à l'aide d'identificateurs sont bien sûr les plus problématiques, mais l'étude que nous avons entreprise a révélé que la proportion de correspondances exactes était étonnamment élevée lorsque l'on procédait à des appariements statistiques, notamment dans le cas des fichiers qui tenaient compte de la structure des ménages.

25. Les ensembles de données raccordés peuvent résulter de:

- a) L'appariement d'ensembles de données de l'ABS;
- b) L'appariement d'un ensemble de données de l'ABS avec un ensemble de données extérieur; ou
- c) L'appariement de deux ensembles de données extérieurs.

Dans les cas a) et b), les ensembles de données obtenus doivent être conservés à l'ABS et pour avoir accès à ces informations, il faut utiliser l'un des moyens de diffusion considérés dans le présent document. Dans le cas c), il n'est pas nécessaire de confier la garde des données à l'ABS, mais il y a des avantages à le faire. Nous sommes en effet régis par des dispositions législatives sur lesquelles pourraient être fondées les modalités d'accès à ces ensembles de données et

la protection de leur confidentialité. En outre, notre réputation est telle que le public sait qu'il peut avoir pleinement confiance en notre capacité de protéger la confidentialité des données. Nous disposons en outre des instruments et des systèmes nécessaires pour permettre l'accès aux données.

26. Un ensemble de données raccordé peut présenter un intérêt analytique considérable comme en témoignent les exemples d'utilisations ci-après:

- Étude des rapports d'une personne avec différentes institutions – en mettant en correspondance, par exemple, des informations sur les soins de santé assurés par les médecins traitants, les hôpitaux, les maisons de retraite, etc.;
- Étude des relations entre les entrées, les sorties et les résultats par la confrontation d'informations ayant trait à la police, aux tribunaux et aux prisons; et
- Étude des évolutions chronologiques en constituant une base de données longitudinales.

27. Quelques principes complémentaires sont nécessaires pour créer ces ensembles de données et les exploiter. Les plus essentiels sont énumérés ci-après.

- La compatibilité avec la mission de l'ABS consistant à utiliser l'information statistique pour favoriser la recherche, le débat et la prise de décisions en toute connaissance de cause.
- Un avantage statistique manifeste.
- L'intégrité et l'ouverture en ce qui concerne les applications.
- La publication d'un produit statistique à partir de chaque ensemble de données raccordées.
- La volonté de conserver la confiance du public en veillant au respect des textes législatifs régissant les activités de l'ABS, la protection des données personnelles et d'autres aspects pertinents.

28. Nous envisageons de créer un comité d'éthique qui nous aidera à prendre des décisions dans ce domaine.

## **V. NOS PROJETS ACTUELS**

29. Jusqu'à une date récente, la situation en ce qui concerne chacun des modes de diffusion des données communiquées aux chercheurs extérieurs pour les aider dans leurs travaux était la suivante:

- Produits statistiques standard – Un service standard était disponible pour tous les domaines de la statistique.
- Cubes de données – Service en cours de mise au point mais déjà disponible pour certaines séries statistiques (données démographiques et données sur la population active, par exemple).

- Services spéciaux de communication des données – Service disponible mais utilisé uniquement à titre occasionnel, généralement pour la production de tableaux détaillés.
- CURF (fichiers d'enregistrements unitaires confidentialisés) – Ce mode de diffusion est employé de manière habituelle pour les enquêtes auprès des ménages, et occasionnelle pour les enquêtes auprès des entreprises, mais il convient de restreindre le degré de détail des données divulguées à ce titre à cause des risques croissants d'appariement. Il convient également de renforcer les engagements juridiques qui sont nécessaires pour la publication des données.
- RADL (Laboratoire de données avec accès à distance) – Service non disponible.
- Laboratoire de données sur le site de l'ABS – Est utilisé occasionnellement mais on n'en encourage pas le développement.
- Collaboration – Cette formule n'a été utilisée qu'occasionnellement par le passé mais au cours de la dernière année environ, près de 40 activités de collaboration ont été mises sur pied dont 12 s'inscrivent dans le cadre du Programme australien d'analyse des données du recensement.
- Analyse interne – Rarement utilisée.

30. À l'avenir, les huit modes de diffusion considérés seront tous utilisés pour soutenir le travail des chercheurs extérieurs. En raison de leur coût, nous nous efforcerons de limiter les services spéciaux de communication des données (troisième mode de diffusion) à quelques clients privilégiés. À cause du risque accru d'appariement, nous procéderons à une certaine réduction de la quantité de détails divulguée au titre des fichiers d'enregistrements unitaires confidentialisés, qui demeureront néanmoins pour les chercheurs un moyen essentiel d'accès aux microdonnées.

31. Nos efforts de développement seront concentrés sur les produits statistiques standard (mode de diffusion 1), le laboratoire de données avec accès à distance (mode de diffusion 5) et la collaboration (mode de diffusion 7). Nous comptons mettre à la disposition des chercheurs un plus grand nombre de cubes de données (mode de diffusion 2) mais pour être tout à fait réaliste, il faut bien admettre que ce type de produit ne convient que pour une gamme limitée de séries statistiques.

32. Notre objectif au titre des «produits statistiques standard» consistera à augmenter le volume de données disponibles sous cette forme par le biais de nos services spéciaux basés sur le Web (AUSSTATS, par exemple). Les responsables de tous les secteurs de la statistique seront priés de revoir leur stratégie de diffusion dans le but de limiter les publications sur papier et d'élargir la gamme des produits électroniques disponibles.

33. Le laboratoire de données avec accès à distance est une option nouvelle qui viendra tout juste d'entrer en service au moment de la tenue de la réunion plénière de la Conférence. Nous estimons que ce mode de diffusion a de l'avenir eu égard aux résultats auxquels a abouti la mise en œuvre de la première version. Ce service visera plus particulièrement à:

- Assurer l'accès aux microdonnées que renferment les ensembles de données plus détaillés; et
- Permettre l'accès à des ensembles de données raccordés, en particulier lorsqu'un de ces ensembles de données est conservé en dehors de l'ABS.

34. À présent que nous disposons d'un Service des analyses pleinement opérationnel et très estimé, nous menons une «collaboration» plus active. Nous nous efforcerons d'instaurer une collaboration dans les domaines qui nous intéressent tout particulièrement, notamment lorsqu'un nouveau produit statistique pourrait en résulter. Nous sommes spécialement intéressés par la possibilité de valoriser les ensembles de données existants au moyen de méthodes d'analyse. Dans la pratique, quelques chercheurs nous contacteront dans un premier temps. Nous déterminerons si une collaboration pourrait s'avérer mutuellement profitable. Le mode de diffusion 8 pourrait se prêter à certains projets de collaboration mais c'est une démarche que nous suivrons de manière sélective.

## **VI. MODALITÉS D'ORGANISATION**

35. L'initiative de ces modalités doit venir de la direction de l'ABS, qui traverse d'ailleurs une période de profondes mutations. La communication est importante, au niveau tant interne qu'externe. Nous bénéficions à cet égard du soutien du service de l'ABS chargé de la politique générale et de la coordination.

36. La gestion et l'administration effectives des modalités d'organisation incombent à notre Division des services d'information. Il existe, au sein de cette division, une unité administrative responsable de la gestion et de la distribution des informations relevant des CURF, du RADL et des laboratoires de données accessibles sur le site de l'ABS. Elle est chargée en outre de la promotion de ces services et des relations avec les clients.

37. Les méthodes d'accès sont encore, à bien des égards, en cours de perfectionnement. Pour renforcer nos moyens de recherche, s'agissant notamment des recherches effectuées à l'extérieur et pour mieux cibler nos efforts, nous avons créé, au sein de la Division de la méthodologie, une unité chargée des méthodes d'accès aux données et de protection de la confidentialité, dirigée par un méthodologiste de haut niveau.

38. Une équipe spéciale de projet (placée sous la surveillance d'un Conseil des projets) a été créée pour appuyer la mise en place du laboratoire de données avec accès à distance. La responsabilité courante de la tenue à jour de ce système est désormais confiée à la Division des services d'information.

39. Le Service des analyses est chargé de mettre sur pied et gérer la plupart des activités de collaboration qui reposent sur l'accès aux microdonnées. Certaines d'entre elles peuvent être gérées par l'intermédiaire du secteur de la statistique concerné mais cela restera l'exception. Nous faisons appel au Service des analyses pour assurer une plus grande cohérence de la démarche suivie. En outre, son personnel possède le savoir-faire nécessaire pour travailler le plus efficacement possible avec nos collaborateurs appartenant au secteur de la recherche. Un Conseil des projets composé de nos statisticiens les plus chevronnés des différentes disciplines statistiques supervise ce travail.

40. Enfin, les spécialistes appartenant aux divers secteurs de la statistique doivent être étroitement associés aux activités considérées. Ce sont eux qui fournissent les ensembles de données de base pour tous les modes de diffusion. Par ailleurs, les chercheurs auront besoin de faire appel de temps à autre à leurs connaissances spécialisées.

## VII. QUESTIONS CLEFS

41. Il devient de plus en plus difficile de fournir des données qui soient véritablement «sécurisées». Par conséquent, nous devons inévitablement nous appuyer davantage sur «une sécurisation du contexte», notamment sur des dispositions juridiques, pour pouvoir apporter un soutien à l'analyse des données secondaire. Ce mode de sécurisation exige un plus grand apport de main-d'œuvre et, partant, des engagements de ressources supplémentaires, alors que les services nationaux de statistique sont souvent confrontés à un manque de ressources. Néanmoins, il nous semble judicieux de réaffecter les ressources nécessaires à cet effet dans une optique d'utilisation efficace de nos données.

42. Encore faut-il que les chercheurs acceptent les dispositions en question. Ils peuvent en effet considérer que celles-ci s'accompagnent de contraintes ou d'une gêne dont ils pourraient fort bien se passer. Ils se demandent pourquoi nous doutons qu'ils agiront comme il se doit? À cet égard, la stratégie de communication est primordiale. Nous devons non seulement informer les chercheurs de ces modalités nouvelles mais aussi leur expliquer pourquoi les contraintes sont nécessaires. Il y a beaucoup plus de chance qu'ils travaillent dans le cadre qui leur est imposé s'ils en comprennent la raison d'être.

43. En réalité, nous sommes en train de passer d'un système qui tend à éviter le risque à une stratégie de gestion du risque. Nous nous exposons à un plus grand risque de perdre la confiance de la population quant au degré de protection de la confidentialité dont nous entourons les données personnelles. Ce risque n'est peut-être que limité et pleinement justifié par la valeur qui s'ajoute ainsi à nos données statistiques, mais il n'en demeure pas moins présent. Les chercheurs n'ont pas les mêmes critères de valeur que les spécialistes de la statistique officielle. À leurs yeux, se sont les impératifs de la recherche qui dominent et ils peuvent se sentir frustrés face à ce qu'ils considèrent comme des obstacles bureaucratiques inutiles. Il est donc inévitable que certains d'entre eux «fassent une entorse aux règles». Il est peu probable qu'un chercheur tente d'identifier une personne en particulier car ce n'est pas ce qui l'intéresse. D'après notre expérience, il est plus probable que les chercheurs enfreignent les règles pour faire progresser leur programme de recherches (nous avons déploré, par exemple, que nos microdonnées étaient parfois revendues pour d'autres recherches, quand bien même il en résultait une valeur ajoutée). Il est important que nous réagissions quand cela se produit. Des sanctions pénales peuvent être appropriées dans certaines circonstances mais il peut s'avérer long et difficile de les imposer. En revanche, il est facile de refuser de fournir des services ultérieurs, notamment à l'institution hôte du chercheur, et cette forme de sanction est très efficace, en particulier si la nouvelle se propage au sein de la communauté des chercheurs que l'ABS n'hésite pas à prendre de telles mesures.

44. On ne doit pas oublier, enfin, qu'il existe une collaboration internationale très active au sein de la communauté des chercheurs. Notre expérience personnelle nous apprend que les chercheurs mettent en avant ce qu'ils peuvent faire dans le pays A par rapport au pays B. Dans cette optique, il serait très utile que nous harmonisions davantage nos approches.

Nous avons arrêté d'un commun accord les Principes fondamentaux de la statistique officielle – pourquoi ne pas adopter des principes fondamentaux relatifs à l'utilisation des microdonnées par les chercheurs extérieurs? Je développerai cette suggestion dans la section suivante.

## VIII. CONCLUSIONS

45. La promotion de l'utilisation de nos données statistiques par des chercheurs extérieurs est un important moyen de conférer davantage de valeur à nos activités statistiques et, de ce fait, elle constitue, à notre avis, l'un des objectifs essentiels de l'ABS. En outre, notre législation nous habilite à appuyer ce genre d'activité.

46. Par le passé, nous avons interprété cette législation de manière conservatrice en privilégiant une démarche qui permettait d'obtenir des données sécurisées. La complexité croissante de la technologie et la disponibilité de bases de données extérieures rendent plus difficile la publication de microdonnées réellement «sécurisées» (ou, du reste, de cubes de données sécurisés). Le rôle grandissant des bases de données du secteur privé est sans doute l'évolution la plus préoccupante car celles-ci sont généralement moins réglementées par des dispositions qui limitent leur utilisation ou préviennent un emploi abusif de leur contenu.

47. Par conséquent, il nous faut évoluer vers une stratégie de diffusion s'appuyant sur un «contexte sécurisé». Nous sommes aidés à cet égard par la confirmation du fait que les engagements juridiquement contraignants signés par les chercheurs entrent en ligne de compte pour déterminer si nous nous conformons à notre législation d'habilitation. En d'autres termes, nous ne sommes plus tenus de nous appuyer uniquement sur des données sécurisées.

48. Nous continuerons d'interpréter avec une assez grande prudence les textes juridiques qui nous autorisent à diffuser des microdonnées, même si nous adoptons une approche un peu moins conservatrice qu'auparavant. Nous craignons, en effet, qu'un seul incident lourd de conséquences puisse nuire gravement à notre réputation et saper la confiance que nous inspirons à la population s'agissant du degré de protection dont nous entourons la confidentialité des données personnelles. Il en résulterait une diminution du taux de réponse lors de nos collectes de données, ainsi qu'une baisse de la qualité des statistiques que nous produisons.

49. La solution nouvelle la plus encourageante pour assurer l'accès aux microdonnées est le RADL (laboratoire de données avec accès à distance), qui présente un intérêt particulier pour développer l'accès aux ensembles de données raccordés.

50. Comme dans de nombreux autres domaines de la statistique, il nous est loisible de profiter mutuellement de nos expériences respectives – bonnes ou mauvaises. L'accès aux microdonnées est sans doute un secteur dans lequel il serait opportun de convenir de principes fondamentaux. Les communautés de chercheurs travaillent au-delà des frontières nationales et peuvent donc procéder à des comparaisons. En réalité, il existe déjà des expériences (voir par exemple l'Étude de Luxembourg sur les revenus) dans le cadre desquelles des microdonnées provenant de plusieurs pays sont rassemblées pour faciliter leur utilisation aux fins de la recherche.

51. Bien évidemment, les dispositions juridiques et administratives en vigueur varient d'un pays à l'autre mais cette diversité ne doit pas nous empêcher d'arrêter d'un commun accord

quelques principes fondamentaux. À titre de point de départ, je suggère de prendre en considération les éléments ci-après:

- Il est souhaitable que les microdonnées recueillies aux fins de l'établissement des statistiques officielles soient exploitées aux fins de la recherche et des analyses des données secondaires, moyennant le respect de certaines prescriptions visant à empêcher qu'il en soit fait mauvais usage.
- Il ne faut pas encourager l'utilisation des microdonnées à d'autres fins extérieures que la recherche et l'établissement de statistiques.
- Il faut que l'utilisation des microdonnées soit régie par des dispositions légales ou autres pour que la population soit convaincue que ces microdonnées sont employées à bon escient.
- Les utilisations des microdonnées devraient être transparentes, et rendues publiques, toujours pour que la population sache qu'il n'en est pas fait un mauvais usage.
- Le service national de statistique ne devrait engager des chercheurs extérieurs en qualité d'employés que s'ils contribuent à l'exécution de travaux qui se concrétisent par un produit de la statistique nationale.
- Les dispositions régissant l'accès aux microdonnées devraient être approuvées par les autorités nationales chargées de la protection de la vie privée.

-----