



Conseil économique et social

Distr. générale
18 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 a) de l'ordre du jour provisoire

**Difficultés rencontrées pour donner accès aux microdonnées
à des fins de recherche: éléments moteurs de l'accès
aux microdonnées**

Échanges de microdonnées et problèmes posés par les données en libre accès et la transparence

**Note établie par le Royaume-Uni et l'Organisation
de coopération et de développement économiques**

Résumé

Le présent document aborde les problèmes posés par les données en libre accès en matière d'échange de microdonnées. Pour servir la communauté des chercheurs, de nombreux bureaux de statistique ont créé des archives nationales de données, des laboratoires de données sur site et des systèmes d'accès à distance. Toutefois, répondre aux besoins du grand public et des entrepreneurs de la nouvelle industrie de l'information présente une série d'autres difficultés. Ceux qui produisent des statistiques officielles devraient tenir compte des contraintes que fait surgir l'anonymisation des données en libre accès. Les bureaux auront besoin d'une gamme de solutions, y compris des laboratoires et des systèmes d'accès à distance, des fichiers réservés aux utilisations scientifiques, des fichiers à usage public et des produits en libre accès. Les services à l'intention de la communauté des chercheurs se sont considérablement améliorés au cours de la dernière décennie, et cette amélioration se poursuivra. Cependant, la création de fichiers à usage public aux normes des données en libre accès destinés à alimenter le développement des données massives (*Big Data*) est beaucoup moins avancée. Ce document examine ces questions, qui pourraient devenir un domaine d'activité important dans les années à venir.

I. Introduction

1. Le présent document se fonde sur l'expérience de l'Office for National Statistics (Bureau national de statistique) du Royaume-Uni, mais n'est pas une déclaration de la politique suivie par le Bureau ou par le Gouvernement britannique. Il aborde quelques-unes des initiatives politiques et législatives concernant le libre accès aux microdonnées dans les bureaux nationaux de statistique. Les éléments moteurs et les solutions en matière de liberté d'accès aux microdonnées sont différents de ceux en matière d'échange de microdonnées confidentielles, mais les problèmes propres à chacune des deux activités ne sauraient être résolus sans une vision claire de l'une et de l'autre.

2. Aux fins du présent rapport, l'expression «données en libre accès» désigne des données (ou des ensembles de données) qui sont:

- a) Accessibles à tout le monde, idéalement par l'Internet;
- b) Transcrites dans un format lisible par une machine numérique et permettant l'interopérabilité avec d'autres données;
- c) Disponibles au coût de reproduction voire moins;
- d) Exemptes de restrictions concernant leur utilisation et leur réutilisation.

3. Aux fins du présent rapport:

- a) Le terme «microdonnées» désigne les données qui permettent à une unité déclarante particulière de quantifier des observations ou des faits;
- b) Le terme «statistique» désigne une combinaison de données susceptible de révéler des tendances dans les observations ou les faits.

4. Aux fins du débat concernant les données en libre accès, les microdonnées pertinentes sont les enregistrements particuliers d'observations directes et de faits obtenus au moyen d'instruments d'enquête statistique, de registres ou de sources administratives. Les données ne sont modifiées avant leur diffusion que si cela est nécessaire pour satisfaire à des obligations de confidentialité à l'égard des sujets des observations. On pourrait appeler de telles données des «microdonnées en libre accès». Le but de ce document est d'étudier l'évolution probable des échanges de microdonnées appelés à se fonder de plus en plus sur les principes s'appliquant aux données en libre accès, et les problèmes que cela pose pour ceux qui produisent les statistiques officielles.

II. Motifs en faveur de l'accès aux microdonnées

5. Parmi les motifs en faveur de l'extension du libre accès aux microdonnées, on peut mentionner l'observation de principes scientifiques, l'appui à la démocratie, la stimulation de la croissance économique et sociale, et l'application des législations et des politiques publiques.

A. Respect des principes scientifiques

6. La diffusion en libre accès renforce le statut de la statistique en tant que discipline scientifique.

7. Les Principes fondamentaux de la statistique officielle¹ requièrent que les méthodes de diffusion des statistiques officielles soient déterminées en fonction de *principes scientifiques*. Il est plus facile pour les utilisateurs d'interpréter les données quand elles sont présentées suivant les normes scientifiques.

8. Quels sont ces *principes* et *normes scientifiques* régissant la diffusion et la présentation des données? Les Principes fondamentaux suggèrent qu'ils sont déjà connus et nous sont même familiers.

9. En juin 2012, la Royal Society a publié son rapport intitulé «Science as an Open Enterprise» (La science, une entreprise ouverte)². Dans ce rapport, la science est considérée comme un processus autocorrecteur. Les théories doivent pouvoir être corroborées, invalidées ou améliorées de façon indépendante. Les conclusions et les données qui les étayent doivent être soumises au public le plus large possible pour être développées plus avant.

10. Ceux qui produisent les statistiques officielles doivent se remettre en question pour savoir s'il est possible de corroborer, d'invalider ou d'améliorer de façon indépendante les statistiques qu'ils publient – ce qu'ils avancent au sujet des tendances qui ressortent des observations et des faits. Le libre accès aux données permet à ce principe scientifique de s'appliquer aux statistiques officielles. La publication de la méthodologie ne permet que des remises en question abstraites, à moins que les données utilisées par le producteur de la statistique soient également accessibles à un tiers indépendant. Cela peut dans une certaine mesure être réalisé en utilisant des voies d'accès contrôlées, par exemple dans le cadre d'un examen statistique par les pairs, ou par l'accès aux données à des fins de recherche. Cependant, la transparence maximale n'est atteinte que lorsqu'aucune restriction, sélection ou pression (réelle ou imaginaire) ne pèse sur un examen indépendant des statistiques.

11. Si les données qui sous-tendent l'information scientifique et les statistiques ainsi que les méthodes utilisées ne sont pas librement accessibles, la confiance du public peut en souffrir. Par exemple, la confiance du public dans les publications du Climate Research Unit (Groupe de recherche sur le climat) de l'université d'East Anglia a été compromise lorsque l'accès à certaines données et méthodes a été refusé au public. Deux enquêtes ont eu lieu, et la réaction du Gouvernement a été de recommander la divulgation complète des données brutes ainsi que des programmes informatiques et méthodologies nécessaires pour reproduire les résultats.

«La divulgation des données brutes et d'indications suffisantes sur les programmes informatiques est primordiale pour encourager les gens à remettre en question la science de la manière habituelle, en contestant les travaux existants, ce qui permet de les valider et d'aller de l'avant avec de nouvelles hypothèses.»

- Réaction des autorités au Premier rapport de la session 2010-2012 de la Commission des sciences et de la technologie³

B. Élément indispensable de la démocratie

12. Les Principes fondamentaux de la statistique officielle qualifient les données des statistiques officielles d'«élément indispensable du système d'information d'une société démocratique».

¹ <http://unstats.un.org/unsd/methods/statorg/FP-French.htm>.

² http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.

³ <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/496/496.pdf>.

13. Le Partenariat pour une gouvernance transparente est une initiative visant à améliorer la gouvernance. Fondé en septembre 2011 par huit gouvernements, il compte désormais 55 gouvernements membres. La déclaration de partenariat comprend un engagement à renforcer la disponibilité des renseignements sur les activités gouvernementales. Les partenaires s'engagent à :

«... fournir de manière volontariste des informations de grande valeur, y compris des données brutes, en temps utile et dans des formats que le public puisse facilement localiser, comprendre et utiliser, et des formats facilitant leur réutilisation».

- Déclaration de gouvernement transparent – Partenariat pour une gouvernance transparente⁴

14. En mettant en œuvre les Principes fondamentaux, la plupart des systèmes statistiques nationaux intégreront des objectifs statutaires, notamment celui de favoriser l'examen des politiques publiques. L'objectif principal de la Statistics Authority (Autorité de la statistique) du Royaume-Uni et de sa branche exécutive, l'Office for National Statistics (Bureau national de statistique) est ainsi libellé :

«Le Conseil aura pour objectif de promouvoir et protéger la production et la publication de statistiques officielles d'intérêt collectif, notion qui comprend notamment l'information du public sur les questions économiques et sociales, et l'aide à l'élaboration et l'évaluation des politiques publiques.».

- Loi de 2007 sur les services de statistique et d'enregistrement⁵

C. Favoriser la croissance économique et sociale

15. Dans une étude réalisée pour la Commission européenne⁶, Graham Vickery, son auteur, a estimé que l'ouverture à la réutilisation des informations du secteur public pourrait générer directement ou indirectement des gains économiques de l'ordre de 140 milliards d'euros.

16. La Recommandation de 2008 du Conseil de l'Organisation de coopération et de développement économiques (OCDE) relative à un accès élargi et une exploitation plus efficace concernant les informations du secteur public⁷ comporte plusieurs recommandations pertinentes en matière de liberté d'accès aux microdonnées statistiques, en vue d'accroître «les retombées économiques et sociales ... du fait notamment d'une distribution plus efficiente [de l'information] d'une innovation accrue et du développement de nouveaux usages.».

17. Les pays membres sont invités à en adopter les principes, parmi lesquels les suivants sont pertinents en ce qui concerne les microdonnées statistiques :

- a) *Ouverture*. Maximiser la disponibilité en faisant de la liberté d'accès la règle générale;
- b) *Conditions transparentes de réutilisation*. Fixer des conditions non discriminatoires, en éliminant les accords d'exclusivité et en supprimant les restrictions inutiles;

⁴ <http://www.opengovpartnership.org/d%C3%A9claration-de-gouvernement-transparent>.

⁵ <http://www.legislation.gov.uk/ukpga/2007/18/section/7>.

⁶ http://ec.europa.eu/information_society/policy/psi/index_en.htm.

⁷ <http://www.oecd.org/fr/internetieconomy/40826098.pdf>.

- c) *Listes de ressources.* Dresser des inventaires des données en libre accès et les publier en ligne;
- d) *Qualité et intégrité.* Employer les méthodes optimales de mise au point des données et de protection contre les représentations mensongères;
- e) *Tarifcation.* Mettre à disposition gratuitement, ou tout au plus à prix coûtant;
- f) *Internationalisation.* Faciliter les utilisations transfrontières et l'interopérabilité.

18. Ces recommandations constituent un cadre visant à permettre aux bureaux nationaux de statistique d'évaluer leur degré d'observation des principes de libre accès aux microdonnées.

D. Lois et mesures visant à libérer l'accès aux données

19. Il existe fondamentalement deux types de lois et mesures visant à libérer l'accès aux données. Le premier type consiste en lois et mesures visant à faire obligation au secteur public de *mettre des données à disposition* en libre accès, et le second, en lois et mesures visant à donner aux utilisateurs le droit d'obtenir librement des données du secteur public.

1. Lois et mesures tendant à la mise à libre disposition des données

20. De nombreux bureaux nationaux de statistique sont soumis à des dispositions légales et réglementaires qui permettent d'espérer qu'ils mettront leurs données à la libre disposition du public en vertu d'un régime réglementaire.

21. Par exemple, dans l'Union européenne, le marché de l'information du secteur public a d'abord été réglementé par la Directive concernant la réutilisation des informations du secteur public⁸. Cette directive est en cours de révision afin de renforcer ses dispositions et de l'adapter à l'évolution de l'univers numérique. À proprement parler, elle ne rend pas obligatoire la production de documents à usage public, mais fournit aux États membres, pour adoption dans leur législation, un cadre réglementaire favorable, partant du principe que les informations du secteur public doivent être librement accessibles dans des limites raisonnables. Elle pose par exemple le principe que le revenu total provenant de la fourniture et de l'autorisation de réutilisation des données ne devrait pas dépasser le coût de leur collecte, de leur production et de leur diffusion augmenté d'un retour sur investissement raisonnable.

22. Cette directive a été mise en œuvre dans tous les États membres, dont certains sont allés au-delà de ses exigences tant dans leur législation interne que dans les mesures qu'ils ont prises. Par exemple, le Royaume-Uni a adopté des Principes en matière de données publiques⁹ contraignants pour le secteur public, parmi lesquels figure l'obligation d'encourager activement l'utilisation et la réutilisation des données publiques des ministères.

23. La réglementation du droit d'auteur dans le secteur public britannique (le «Crown copyright») a été modifiée de manière à mettre en vigueur un principe de mise à libre disposition des données. La licence de droit d'auteur par défaut en vigueur dans le secteur public britannique est l'Open Government Licence (licence de libre accès à l'information

⁸ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:FR:NOT>.

⁹ <http://www.data.gov.uk/library/public-data-principles>.

publique)¹⁰, une licence qui accorde de manière positive et volontariste la mise à libre disposition des données:

«[Le Gouvernement] vous accorde une licence mondiale, libre de redevances, perpétuelle et non exclusive d'utilisation de l'information [...]. Vous êtes libre de copier, publier, diffuser et communiquer cette information; de l'adapter; de l'exploiter commercialement, par exemple en la combinant avec d'autres informations, ou en l'intégrant dans votre propre produit ou logiciel.»

- Licence de libre accès à l'information publique du Gouvernement britannique

24. Tout organisme public britannique souhaitant obtenir une dérogation à la règle limitant le prix des données à la récupération de leur coût marginal doit demander l'agrément de l'Information Fair Trader Scheme (Régime pour l'équité dans les échanges d'informations)¹¹ et présenter une demande motivée soumise à l'examen de l'Office for Public Sector Information (Bureau des informations du secteur public).

25. À l'avenir, on peut s'attendre à voir se développer les mesures et les structures tendant à la mise à libre disposition des données, souvent à la suite de campagnes d'opinion¹². Le Royaume-Uni lance actuellement l'Institut des données en libre accès (Open Data Institute, ODI)¹³ sous la forme d'un partenariat public-privé. Coprésidé par Nigel Shadbolt et Tim Berners Lee, l'Institut vise expressément à diffuser les données en libre accès à l'extérieur du secteur public:

«[Il s'agit d'une] collaboration entre nos principaux entrepreneurs et entreprises, universités et chercheurs, le Gouvernement et la société civile pour débloquer la valeur capitalistique et sociale de la vaste quantité de données publiques en libre accès actuellement rendues accessibles.»

- À propos de l'Institut des données en libre accès», www.theODI.com.

2. Lois et mesures tendant à donner aux utilisateurs le droit d'obtenir des données

26. De nombreux bureaux nationaux de statistique sont également soumis à des lois et mesures qui donnent au public le droit d'obtenir des données du secteur public. À titre d'exemple, on peut citer les règlements relatifs à l'information concernant l'environnement (Environmental Information Regulations), et les lois relatives à la liberté de l'information qui existent dans de nombreux pays. Il arrive souvent que les données statistiques soient soumises à ces lois et mesures, et que ne puissent en être exemptées que les données dont la confidentialité peut être démontrée et celles dont la publication est prévue ultérieurement sous la forme demandée. Les conséquences de ces lois sur la diffusion des données statistiques peuvent être importantes:

a) De telles lois fixent généralement un délai d'exécution. Les bureaux nationaux de statistique du Royaume-Uni doivent répondre dans les vingt jours suivant la demande en fournissant soit les données demandées soit une explication des raisons pour lesquelles une exemption s'applique;

b) Leur association avec la licence de libre accès à l'information publique (ou son équivalent) empêche d'imposer des limitations ou des conditions concernant l'utilisation et la réutilisation des données;

¹⁰ <http://www.nationalarchives.gov.uk/doc/open-government-licence>.

¹¹ <http://www.nationalarchives.gov.uk/information-management/ifts.htm>.

¹² <http://www.freeourdata.org.uk>.

¹³ <http://www.theodi.org/people/nrs>.

c) Le demandeur a généralement le droit de faire appel d'une décision de rejet de la totalité ou d'une partie de la demande de données. Cet appel est généralement jugé par un tribunal chargé de l'information ou une autre autorité non statistique. Le bureau de statistique concerné peut présenter les éléments justifiant la non-divulcation des données (par exemple pour des raisons de confidentialité), mais la décision relève du tribunal;

d) En règle générale, il n'est pas nécessaire de justifier la demande de données. Le demandeur n'a pas à expliquer la raison de sa demande, ni l'usage qu'il prévoit de faire des données;

e) Aucune exception n'est en général prévue pour des données de mauvaise qualité;

f) La diffusion de données en vertu de lois tendant à donner un droit d'obtention aux utilisateurs est souvent de nature à créer un précédent;

g) Si elle n'est pas effectuée avec précaution, la diffusion de données en vertu de lois tendant à donner un droit d'obtention aux utilisateurs peut ressembler à l'octroi d'un accès privilégié;

h) La fourniture d'ensembles de données peut porter atteinte à la confidentialité en cas d'attaque en mosaïque;

i) Les questions touchant au contrôle de la divulgation sont souvent obscures pour les non-initiés.

27. Les lois et mesures tendant à la mise à libre disposition des données et celles tendant à donner aux utilisateurs le droit d'obtenir des données permettent désormais aux entrepreneurs de l'information, qu'ils soient analystes financiers ou concepteurs d'applications, d'investir sereinement dans des experts de l'exploitation des données. L'industrie qui en découle exerce à son tour une pression sur les organismes détenant des données du secteur public.

III. Problèmes posés par la production de (micro)données en libre accès

28. Comment les bureaux nationaux de statistique réagissent-ils aux problèmes que pose la liberté d'accès aux données? En ce qui concerne leurs statistiques, on pourrait soutenir que la réaction est très bonne. Les pratiques fondamentales en matière de présentation et de diffusion des statistiques sont par nature compatibles avec le principe de liberté d'accès aux données. Toutefois, lorsque l'on considère les questions relatives à l'accès aux microdonnées brutes des bureaux nationaux de statistique, la réaction est actuellement moins convaincante. En règle générale, un bureau de statistique classe les données qu'il possède en trois catégories:

a) Les «statistiques» destinées à la publication ou à une utilisation dans l'administration publique – il s'agit en général de données ou de visualisations agrégées;

b) Les «séries de données à usage de recherche» destinées à des fins de recherche statistique dans des conditions contrôlées – il s'agit en général de microdonnées et de données habituellement estampillées «confidentiel»;

c) Les «données à usage de production», qui sont des microdonnées ou des données agrégées faisant partie de systèmes de production de statistiques et qu'il n'est pas prévu d'utiliser à des fins autres que le calcul de statistiques et de séries de données à usage de recherche.

29. La notion de «fichiers à usage public» est courante, mais la pratique actuelle consiste généralement à l'affecter à la catégorie des «séries de données à usage de recherche». Cette classification est habituellement correcte, car la plupart des fichiers à usage public ne sauraient être considérés à l'instar des données en libre accès, en ce sens qu'ils sont soumis à des conditions d'utilisation qui peuvent porter sur leur actualité, leur exhaustivité, leur granularité, leur format et leur interopérabilité. Actuellement, ni les bureaux nationaux de statistique ni les utilisateurs de fichiers publics ne s'attendent que ces fichiers puissent servir à reproduire les statistiques produites par les bureaux nationaux de statistique – ne serait-ce que parce que celles-ci ne sont pas produites à partir de ces fichiers. Étant donné la manière dont elles sont actuellement produites, la diffusion des fichiers à usage public ne permet donc pas d'appliquer aux statistiques le principe scientifique de corroboration, de vérification et d'amélioration indépendantes.

30. Il n'est pas courant à l'heure actuelle de considérer que les microdonnées puissent être placées dans la catégorie des données en libre accès et de placer des informations dans cette catégorie en vue de leur diffusion dans les conditions correspondantes. Cela découle peut-être des nombreux problèmes que pose la production de microdonnées en libre accès.

A. Confidentialité

31. Les questions de confidentialité sont le motif le plus fréquemment cité pour ne pas produire de microdonnées en libre accès. Il s'agit très souvent d'un motif tout à fait légitime. Cependant, il est de pratique courante pour un bureau national de statistique de considérer l'ensemble de ses données inédites comme des données confidentielles, soit par défaut, soit en vertu de leur statut de «données détenues aux fins de la statistique officielle». Il est peu probable que cette pratique survive aux deux types de lois et mesures tendant à instaurer un libre accès aux données. Tout d'abord, les politiques des bureaux nationaux de statistique doivent tenir compte du fait que certaines données sont intrinsèquement moins susceptibles que d'autres de poser des problèmes de confidentialité – par exemple, les données sur le budget et les dépenses du secteur public, ou les données sur les prix, peuvent être classées comme microdonnées mais pas comme données confidentielles, les informations correspondantes étant déjà disponibles dans le domaine public. En second lieu, les bureaux nationaux de statistique devraient analyser de manière plus systématique et plus critique leurs ensembles de données pour déterminer si ces ensembles doivent véritablement ou non être protégés. Ces analyses peuvent comporter des «tests d'intrusion», dans lesquels un tiers de confiance se voit fournir un ensemble de données à partir desquelles il tente de découvrir, par des moyens raisonnables, des renseignements d'ordre privé et confidentiel sur des individus déterminés. Si ce tiers peut découvrir de tels renseignements personnels, cela constitue pour le bureau national de statistique un élément de preuve indépendant qu'il peut présenter à un tribunal chargé de l'information ou autre pour étayer, en cas de litige, un refus de fournir les données concernées à un demandeur. Si le test d'intrusion n'aboutit pas à la découverte de renseignements personnels de ce genre, le bureau national de statistique a peut-être découvert un ensemble de microdonnées susceptible d'être placé en libre accès.

32. Les organismes responsables de la protection des données sont conscients qu'il est difficile de distinguer entre les renseignements d'ordre personnel et les autres. L'avis du Groupe de travail «article 29» sur le concept de données à caractère personnel¹⁴ aide les bureaux nationaux de statistique à déterminer ce qui constitue ou non des données personnelles. Le Bureau du Commissaire à l'information (Information Commissioner's Office) du Royaume-Uni a récemment publié, avec l'aide du Bureau national de statistique

¹⁴ http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf.

et de nombreuses autres institutions, un Code de bonnes pratiques sur l'anonymisation des données personnelles et la gestion du risque en matière de protection des données¹⁵.

B. Qualité et réputation des données

33. De tous les facteurs susceptibles d'altérer la confiance du public dans les statistiques officielles, la qualité des données est peut-être le plus important. Pour les bureaux de statistique, il est difficile de se faire une réputation de qualité et facile de la perdre. Il est tout à fait compréhensible qu'ils ressentent le libre accès aux données, et notamment aux microdonnées, comme une menace pour la réputation de qualité de leurs statistiques. Cette menace peut se concrétiser de deux manières. Premièrement, les données peuvent très bien être de mauvaise qualité. Si elles sont exigées au nom du droit du public à l'information, il se peut qu'elles soient communiquées avant que leur qualité ait été vérifiée et que le bureau national de statistique ait pu recevoir des données supplémentaires. Deuxièmement, il est possible que les données soient de bonne qualité mais qu'elles soient interprétées différemment ou même de manière erronée par l'autre partie. Le bureau national de statistique risque de devoir expliquer les différences et les justifier auprès d'un public qui préfère peut-être que l'autre partie ait raison. Pour affronter ces difficultés, il faut disposer d'excellentes métadonnées, être en mesure de commenter la qualité de l'analyse de l'autre partie, et mener auprès du public une campagne d'information sur les questions de qualité et d'analyse. Il est possible de résoudre le problème culturel en s'appuyant sur les bonnes pratiques scientifiques – lorsque les méthodes statistiques sont de bonne qualité, il n'y a rien à craindre de la réutilisation des données de base lorsque celles-ci deviennent en libre accès.

C. Normes de diffusion

34. Les normes s'appliquant à la diffusion des microdonnées en libre accès sont contraignantes. Peu de bureaux nationaux de statistique seront en mesure de s'y conformer à court terme. Le secteur public britannique est confronté au défi que représente la notation en cinq étoiles des données en libre accès¹⁶. En particulier, l'utilisation de formats non propriétaires et d'identifiants uniformes de ressources (URI) pour toutes les microdonnées placées en libre accès représente un pari, et pourrait même être mal accueillie par les utilisateurs actuels.

D. Utilisation de sites Web conçus à d'autres fins

35. Par rapport aux statistiques agrégées, les microdonnées en libre accès peuvent être très volumineuses et être stockées dans un format incompatible avec les moyens de diffusion existants, tels que les sites Web des bureaux nationaux de statistique. Il se peut qu'une révision de la conception de ces sites pour permettre le libre accès aux données ne soit pas prioritaire, surtout s'il existe des grosses contraintes telles que les bandes passantes ou la taille des fichiers, et des restrictions en matière de format propriétaire. Une solution consiste à adopter d'autres modes de diffusion spécialement conçus. Au Royaume-Uni, par exemple, a été mis en place le site Web www.data.gov.uk, qui est un service commun à la disposition de toutes les institutions du secteur public. Il convient également de garder à l'esprit que les restrictions imposées par les normes de sécurité concernant les informations

¹⁵ http://www.ico.gov.uk/news/latest_news/2012/~/-/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx.

¹⁶ <http://5stardata.info>.

gouvernementales ne sont pas applicables aux données en libre accès, ce qui permet d'utiliser les moyens de diffusion du secteur privé. Un exemple en est fourni par Google Public Data, qui accueille actuellement les données en libre accès de l'OCDE, d'Eurostat, du Fonds monétaire international, de la Banque mondiale, du Bureau fédéral de statistique allemand (Destatis) et du Bureau des recensements des États-Unis. Google Public Data est avant tout un agrégateur de données publiées par d'autres voies, mais pourrait être utilisé pour la publication initiale de données placées en libre accès.

E. Authenticité et attribution

36. Les bureaux nationaux de statistique sont la source originale attribuée des microdonnées placées en libre accès, mais peuvent éprouver des préoccupations légitimes au sujet d'éventuelles modifications du contenu des données qui auraient une incidence sur le produit des analyses mais ne seraient pas déclarées ou expliquées par les utilisateurs. En d'autres termes, les données perdraient leur authenticité, alors qu'elles continueraient d'être attribuées à un bureau national de statistique. La solution réside dans l'utilisation, par le producteur de données en libre accès, d'identifiants persistants. Chaque élément de données en libre accès doit être associé à un identifiant d'objet numérique ou un nom uniforme de ressource unique, combiné à un localisateur uniforme de ressource permanent et à une citation unique pour les publications électroniques utilisant la norme ISO 690-2 de l'Organisation internationale de normalisation. Cela permettrait aux bureaux nationaux de statistique et aux utilisateurs de situer et récupérer facilement les données intactes de la source primitive.

F. Métadonnées

37. La norme relative aux métadonnées la plus pertinente pour les statistiques nationales est la norme SDMX (Échange de données et de métadonnées statistiques), mais elle est optimisée pour les données statistiques agrégées plutôt que pour des microdonnées en libre accès. Si un bureau national de statistique ne dispose pas du savoir-faire nécessaire concernant les normes relatives aux métadonnées pour les microdonnées en libre accès, il peut, pour franchir l'obstacle, demander de l'aide à des partenaires tels que les archivistes de données, en particulier aux membres du Conseil européen des archives de données en sciences sociales (CESSDA).

G. Respect des engagements au titre du Code de bonnes pratiques en matière de publications programmées et d'égalité d'accès en cas de conflit de calendrier avec les obligations au titre des normes relatives aux données en libre accès

38. Les bureaux nationaux de statistique voudront s'assurer que les statistiques dont la publication est prévue à date fixe annoncée à l'avance, conformément aux bonnes pratiques en la matière, ne sont pas compromises par des statistiques similaires issues de la réutilisation de microdonnées en libre accès. Lorsque les microdonnées en libre accès sont mises à la disposition du public de manière organisée, il est facile de s'en assurer en programmant les deux publications de manière adéquate. Toutefois, lorsque les microdonnées peuvent être obtenues sur simple demande, il peut y avoir conflit entre leur communication et le calendrier de diffusion du bureau concerné. Une solution pourrait être soit d'avancer la sortie officielle prévue, lorsque cela est possible, soit d'expliquer préventivement que des statistiques non authentiques et non autorisées provenant d'autres sources peuvent être diffusées avant la publication officielle.

39. Lorsque les microdonnées en libre accès sont mises à la disposition des utilisateurs à leur demande par les bureaux nationaux de statistique, il est important de préserver le principe de l'égalité d'accès. Il s'agit, en statistique, d'un principe important, qui s'impose également en matière de données en libre accès. Par exemple, fournir à un demandeur des données ou des microdonnées en libre accès sous forme de pièces jointes dans un courriel pourra ressembler à l'octroi d'un accès privilégié accordé à ce dernier, et l'utilisation des données peut paraître lui avoir été réservée en exclusivité. Les bureaux nationaux de statistique doivent créer sur leur site Web un registre des mises à disposition¹⁷ présentant les communications ponctuelles de données en libre accès demandées par tel ou tel utilisateur mais clairement déclarées par ailleurs accessibles à tous. Un utilisateur ayant obtenu communication des données ne devrait se voir fournir un lien avec l'URL concernée que si ces données sont accessibles à tous.

H. Allocation des ressources

40. Les caractéristiques d'utilisation des ressources en matière de microdonnées en libre accès sont différentes de celles des données à usage de recherche. Même si, par définition, l'information brute existe déjà, les coûts initiaux de ces ressources sont élevés, notamment pour l'analyse du risque que présente la divulgation des données, leur assemblage dans un format ouvert, la création de métadonnées, etc. En principe, les revenus tirés des microdonnées en libre accès ne peuvent pas dépasser les coûts augmentés d'un retour raisonnable sur l'investissement consenti pour leur production. Cependant, une fois ces microdonnées produites, les coûts permanents sont négligeables, en particulier si le moyen de diffusion est une ressource partagée ou est à la charge d'un tiers. En revanche, les caractéristiques d'utilisation des ressources en matière de séries de données à usage de recherche sont au départ limitées, étant donné que les données sont brutes de production. S'agissant de la production de séries de données à usage de recherche, le maintien d'un cadre de recherche sécurisé, l'accréditation des chercheurs et des projets de recherche et la vérification des résultats entraînent des coûts élevés. Les budgets des bureaux nationaux de statistique privilégient en général la couverture des coûts d'activités en cours, pour lesquelles les ressources sont plus faciles à allouer, même si, sur plusieurs années, la charge totale que représentent les séries de données à usage de recherche est supérieure à l'investissement dans des microdonnées en libre accès d'une valeur équivalente pour la société.

IV. Traitement des problèmes – Projets de recommandation

41. Le présent rapport a examiné les éléments moteurs de l'échange de microdonnées et les problèmes que cela pose dans le cadre des nouvelles normes en matière de liberté d'accès aux données.

42. Respecter ces normes en ce qui concerne les microdonnées officielles présente manifestement des avantages très importants:

a) Le placement des microdonnées en libre accès permet de respecter les *principes scientifiques* de corroboration, de validation et d'amélioration des statistiques officielles provenant des mêmes sources;

¹⁷ Voir <http://www.ons.gov.uk/ons/about-ons/what-we-do/FOI/foi-requests/index.html>.

b) La liberté d'accès aux microdonnées permet de les échanger sans obstacles administratifs coûteux et bureaucratiques;

c) Les microdonnées en libre accès peuvent être utilisées à toutes fins, y compris celles qui n'avaient pas été envisagées lors de leur production;

d) La liberté d'accès aux microdonnées permet aux bureaux nationaux de statistique et aux tiers de tous les secteurs de coopérer et collaborer pleinement autour d'une ressource commune;

e) Les microdonnées en libre accès constituent une catégorie supplémentaire de ressources en information pour les bureaux nationaux de statistique qui, de ce fait, devraient à l'avenir devenir des partenaires importants et (espérons-le) précieux du secteur public dans les sociétés de l'information modernes;

f) La liberté d'accès aux microdonnées favorise le développement des entreprises de l'information et stimule la croissance économique et sociale.

43. Les obstacles et les problèmes sont eux aussi considérables:

a) Les risques pour la confidentialité et la notion même de confidentialité doivent être pris en compte;

b) Il peut se poser des problèmes logistiques qui remettent en question l'architecture d'entreprise des bureaux nationaux de statistique;

c) L'authenticité et l'identification des ressources doivent être prises en compte;

d) Les normes qui s'appliquent aux données en libre accès sont contraignantes.

A. Recommandations

- Les bureaux nationaux de statistique devraient adopter les normes relatives aux données en libre accès pour leur production statistique ordinaire. Ils se familiariseraient ainsi avec les problèmes liés à ce libre accès avant de s'attaquer aux problèmes particuliers que pose le libre accès aux microdonnées;
- Les bureaux nationaux de statistique devraient prévoir une catégorie dédiée aux microdonnées en libre accès dans leurs registres de ressources en matière d'information;
- Les bureaux nationaux de statistique devraient collaborer de manière à répartir les coûts inhérents à l'élaboration de méthodologies pour la création de microdonnées en libre accès;
- Les bureaux nationaux de statistique devraient étudier des voies de diffusion de substitution pour les microdonnées en libre accès si leur architecture n'est pas adaptée;
- Les bureaux nationaux de statistique devraient collaborer étroitement avec leur organisme national de tutelle chargé de la protection des données en vue d'établir des normes d'anonymisation et de préciser la notion de données à caractère personnel;
- Les bureaux nationaux de statistique devraient étudier à l'avance comment ils pourront respecter les obligations qui leur incombent en vertu des codes de bonnes pratiques en matière de statistique lorsqu'ils produiront des microdonnées en libre accès;

- Les bureaux nationaux de statistique devraient utiliser les compétences et l'expérience des conservateurs de données, des informaticiens, des professionnels de la gestion des connaissances et de l'information, des archivistes de données et des bibliothécaires nationaux pour contribuer à l'élaboration et la diffusion de microdonnées en libre accès;
 - Les bureaux nationaux de statistique devraient élaborer une stratégie de communication auprès des principaux utilisateurs de statistiques et du grand public concernant les nouveaux produits que sont les données en libre accès.
-