



Conseil économique et social

Distr. générale
20 décembre 2013
Français
Original : anglais

Commission de statistique

Quarante-cinquième session

4-7 mars 2014

Point 3 j) de l'ordre du jour provisoire*

Questions soumises à la Commission pour examen

et décision : traitement massif des données

informatiques et modernisation

des systèmes de statistique

Traitement massif des données informatiques et modernisation des systèmes de statistique

Rapport du Secrétaire général

Résumé

Établi conformément à la décision 2013/235 du Conseil économique et social, le présent rapport rend compte de l'utilisation des mégadonnées (big data) dans les statistiques officielles à l'heure actuelle. Il donne une vue d'ensemble des dernières initiatives dans ce domaine et expose les conclusions de l'étude que la Division de statistique a consacrée au sujet en 2013.

Ouvrant sur l'avenir, il préconise la création d'un groupe de travail d'envergure mondiale qui s'inspirera de l'expérience acquise à l'occasion d'initiatives régionales favorisant les échanges en ce qui concerne les innovations méthodologiques, les meilleures pratiques s'agissant des questions stratégiques et les possibilités de formation. Ce groupe aura également pour mission de faciliter les partenariats internationaux en vue de l'utilisation des mégadonnées dans le cadre des transferts de technologies aux pays en développement et à l'appui du programme de développement pour l'après-2015. La Commission de statistique est invitée à faire connaître son point de vue sur les propositions figurant dans la dernière partie du rapport.

* E/CN.3/2014/1.



I. Introduction

1. Le séminaire du vendredi, qui était consacré aux questions nouvelles, avait pour thème les mégadonnées au service de la prise de décisions, du développement et des statistiques officielles¹. Des représentants du secteur privé et des responsables des services de statistique se sont exprimés sur l'utilité des mégadonnées pour l'élaboration des politiques et sur le rôle que devraient jouer les systèmes statistiques nationaux dans l'exploitation de ces nouvelles sources d'information. Présidée par le responsable de l'institut de statistique de l'Australie, la séance de l'après-midi a mis en évidence le fait que les mégadonnées constituaient, pour les statisticiens des organismes officiels, une source d'information incontournable, et que la profession devait s'organiser et prendre de toute urgence les mesures voulues pour en tirer partie tout en s'attaquant aux problèmes qui se posaient.

2. À la quarante-quatrième session de la Commission de statistique, à titre de suivi immédiat, le représentant de l'Australie a demandé qu'une étude sur l'exploitation des mégadonnées dans les statistiques officielles soit présentée à la Commission à sa quarante-cinquième session. Le présent rapport fait suite à cette demande.

3. L'utilisation généralisée des outils électroniques ainsi que la production massive d'informations numériques mettent aujourd'hui à la disposition des statisticiens un flux abondant et permanent de données d'un type radicalement nouveau : les mégadonnées. Très dispersées, peu structurées, abondantes et souvent disponibles en temps réel, ces données diffèrent considérablement des données classiques. Sources d'information abondantes, dynamiques et variées, les mégadonnées doivent être traitées selon des méthodes coûteuses et innovantes pour enrichir l'analyse et permettre de prendre des décisions plus avisées.

4. Alors que le taux de réponse aux enquêtes nationales auprès des ménages et des entreprises est en diminution, les mégadonnées peuvent fournir aux décideurs des informations en temps réel, notamment sur les prix, l'emploi, la production et le développement économiques et la démographie². Elles peuvent même permettre d'établir des statistiques plus pertinentes et plus actuelles que les sources traditionnelles de statistiques officielles, comme les enquêtes et les données administratives. La plupart des mégadonnées proviennent du secteur privé, et la plupart des pays n'ont pas encore légiféré sur les conditions de leur exploitation pour la production de statistiques officielles. Il convient donc de continuer d'examiner avec tous les acteurs concernés les questions juridiques et les problèmes de confidentialité et de méfiance à l'égard des mégadonnées, ainsi que la voie à suivre pour les exploiter de manière fiable et crédible en temps opportun. En intégrant les mégadonnées dans les statistiques officielles, les instituts de statistique nationaux, régionaux et internationaux pourraient obtenir des informations sur

¹ Ce séminaire s'est tenu le vendredi 22 février 2013, en marge de la quarante-quatrième session de la Commission de statistique. On trouvera des précisions à ce sujet, en anglais uniquement, à l'adresse suivante : http://unstats.un.org/unsd/statcom/statcom_2013/seminars/Big_Data/default.html.

² Organisation de coopération et de développement économiques (OCDE), « Exploring data-driven innovation as a new source of growth: mapping the policy issues raised by big data » (DSTI/ICCP(2012)9/FINAL), 2013 (*en anglais uniquement*).

l'économie, la société et l'environnement plus rapidement, à moindre coût et en déployant moins de moyens.

5. Prenant acte des innovations technologiques et de l'explosion du nombre et de la diversité d'informations disponibles en temps réel, le Secrétaire général de l'Organisation des Nations Unies a lancé, en 2009, l'initiative Global Pulse, dont l'objectif est de tirer parti des innovations en matière de données numériques, de collecte rapide de données et d'analyse, pour permettre aux décideurs de se faire une idée en temps réel de la façon dont les crises affectent les populations vulnérables. Il est indiqué dans le livre blanc publié dans le cadre de cette initiative que si les mégadonnées ne sont pas la panacée, elles nous offrent une occasion sans précédent d'améliorer notre capacité commune d'aider et de protéger les populations en étudiant les volumes d'informations numériques de plus en plus importants qu'elles produisent³.

6. Parmi les recommandations que le Groupe de personnalités de haut niveau chargé d'étudier le programme de développement pour l'après-2015 a faites au Secrétaire général dans son rapport intitulé « Pour un nouveau partenariat mondial : vers l'éradication de la pauvreté et la transformation des économies par le biais du développement durable », figure un appel à une « révolution en matière de données ». Les mégadonnées, que l'on considère comme un volet important de cette révolution, permettraient d'améliorer la qualité des statistiques, notamment leur actualité et leur exhaustivité, sans compromettre leur pertinence, leur impartialité et leur rigueur méthodologique. Dans son rapport, le Groupe de personnalités de haut niveau estime que si les États disposaient de meilleures informations et statistiques, ils seraient davantage en mesure de suivre les progrès réalisés et de fonder leurs décisions sur les faits, ainsi que de préciser les responsabilités de chacun. Cela ne concerne pas seulement les États, mais aussi les organismes internationaux, les organisations de la société civile et le secteur privé. Une véritable révolution des données devrait s'appuyer autant sur les sources de données existantes que sur les nouvelles pour intégrer pleinement les statistiques dans la prise de décisions, favoriser le plein accès aux données et leur libre utilisation et renforcer les systèmes statistiques » (chap. 4, p. 24).

7. À une échelle plus large, on sait que les mégadonnées sont susceptibles de changer maints aspects de la société. La Commission des mégadonnées de la fondation TechAmerica a estimé, dans un rapport récent, que les mégadonnées pouvaient transformer jusqu'à l'État et la société⁴. Les données d'aujourd'hui, de par leur surabondance, leur variété et leur vitesse d'évolution, recelaient des informations, des faits, des relations, des indicateurs et des indices nouveaux, qui soit ne pouvaient pas matériellement être découverts par le passé, soit n'existaient tout simplement pas⁵. Collectées, gérées et analysées efficacement, ces nouvelles informations pouvaient considérablement renforcer l'efficacité de l'État⁶.

8. La section II du présent rapport donne une vue d'ensemble des activités relatives aux mégadonnées et à la modernisation des systèmes de statistique. La

³ Global Pulse, « Big data for development: challenges and opportunities », mai 2012.

⁴ Phrase liminaire du résumé du rapport de la Commission des mégadonnées de la fondation TechAmerica, intitulé « Demystifying big data: a practical guide to transforming the business of government », Washington, octobre 2012 (*en anglais uniquement*).

⁵ Ibid., deuxième phrase.

⁶ Ibid., troisième phrase.

section III récapitule les différentes sources de mégadonnées et les problèmes qu'elles posent. La section IV présente les résultats de l'étude mondiale sur l'exploitation des mégadonnées dans la production de statistiques officielles que la Division de statistique a menée de juillet à octobre 2013. La section V propose quelques études de cas sur l'exploitation des mégadonnées. Le rapport s'achève sur une proposition d'orientation qui s'inscrit dans le droit fil des initiatives des organisations partenaires et met l'accent sur les problèmes relatifs à l'emploi des métadonnées pour l'établissement de statistiques officielles par les pays en développement et pour le programme de développement pour l'après -2015.

II. Aperçu général des activités récentes

9. Conscients de l'importance du phénomène, un nombre croissant de pays ont mis sur pied des réseaux de professionnels et des groupes de travail pour réfléchir à l'exploitation des mégadonnées et à ses retombées éventuelles. La profession est en train de se rendre compte qu'un changement de paradigme est probablement en cours. L'Institut national italien de statistique reconnaît que, pour exploiter les trésors, que l'on devine immenses sous la montagne des mégadonnées, les instituts nationaux de statistique devant pas avoir le vertige⁷. L'Office central de la statistique des Pays-Bas estime qu'en ce qui concerne l'exploration des potentialités des mégadonnées, les instituts officiels de statistique ont à peine effleuré la surface⁸.

10. La Commission économique pour l'Europe (CEE), EUROSTAT, l'Organisation de coopération et de développement économiques (OCDE) et la Commission économique et sociale pour l'Asie et le Pacifique (CESAP) ont organisé, du 23 au 25 avril 2013, une réunion sur la gestion des systèmes d'information statistique qui s'est tenue simultanément à Paris et à Bangkok⁹. La principale conclusion de cette manifestation a été que les instituts de statistique devaient unir leurs efforts pour régler les problèmes auxquels ils sont tous confrontés face à l'exploitation des mégadonnées et s'efforcer de trouver des solutions communes, notamment en produisant des statistiques à espérance de vie réduite, en adoptant une démarche pluridisciplinaire et en convenant ensemble d'une classification commune des différents types de mégadonnées.

11. Les participants au cinquante-neuvième Congrès mondial de la statistique de l'Institut international de statistique (IIS), qui s'est tenu dans la Région administrative spéciale de Hong Kong (Chine) entre le 25 et le 30 août 2013, se sont longuement penchés sur la question des mégadonnées. Le programme de cette manifestation comprenait un exposé sur le sujet, une séance thématique spéciale sur les possibilités offertes, dans les domaines des statistiques officielles, par Internet, les mégadonnées et les données organiques, et une manifestation connexe consacrée aux mégadonnées et au calcul statistique. Lors de la séance thématique spéciale,

⁷ Monica Scannapieco, Antonino Virgillito et Diego Zardetto, « Placing big data in official statistics: a big challenge? », 21 décembre 2012 (*en anglais uniquement*).

⁸ Piet Daas et Mark van der Loo, « Big data (and official statistics) », document de travail élaboré en vue de la réunion de 2013 sur la gestion des systèmes d'information statistique, Paris et Bangkok, 23 au 25 avril 2013.

⁹ Disponible à l'adresse suivante (*en anglais uniquement*) : <http://www1.unece.org/stat/platform/display/msis/MSIS+2013>.

John Dunne, de l'Office central irlandais de statistique, a présenté une étude intitulée « Big Data coming soon ... to a National Statistical Institute near you » (Les mégadonnées ... bientôt dans votre Institut national de statistique)¹⁰ décrivant certaines sources de mégadonnées telles que les informations relatives à l'itinérance en téléphonie mobile, la consommation électrique ou les paiements électroniques, et montrant comment ces grandes quantités d'informations fournies en temps réel pouvaient être traitées par des prestataires extérieurs grâce à l'informatique en nuage. Les grands ensembles de données seront également l'un des thèmes principaux du programme scientifique du soixantième Congrès mondial de la statistique, qui doit se tenir à Rio de Janeiro (Brésil) du 26 au 31 juillet 2015¹¹. Les congrès de l'Institut international de statistique favorisent la coopération et les échanges de connaissances entre statisticiens issus des instituts de statistique, des milieux universitaires et du monde des affaires.

12. EUROSTAT étudie actuellement les possibilités qu'offre l'exploitation des mégadonnées dans les statistiques officielles portant, par exemple, sur les prix (à partir des données présentes sur Internet) et sur l'utilisation des technologies de l'information et des communications (TIC). D'autre part, lors de la réunion annuelle des directeurs généraux des instituts nationaux de statistique européens, qui s'est tenue en septembre 2013 à La Haye, une session a été consacrée aux mégadonnées¹². Ces travaux ont permis d'élaborer le Mémoire de Scheveningen sur les mégadonnées et les statistiques officielles (adopté, le 27 septembre 2013, par le Comité du système statistique européen), qui encourage les membres du système statistique européen à mettre au point une stratégie en rapport aux mégadonnées, à faire part de leur expérience et à collaborer à l'échelle du système statistique européen et à un niveau supérieur. Un plan d'action et une feuille de route devraient être adoptés d'ici à la mi-2014 et intégrés au programme de travail d'EUROSTAT.

13. Les mégadonnées figuraient également au programme des débats du séminaire que la Commission économique pour l'Europe a consacré à la collecte des données statistiques (Genève, 25 au 27 septembre 2013)¹³. À l'occasion de cette manifestation, MM. Peter Struijs et Piet Daas, de l'Office central de la statistique des Pays-Bas, ont présenté un document de travail (n° 31) intitulé « Big Data, big impact? », dans lequel ils avançaient que les mégadonnées allaient certainement avoir d'importantes conséquences pour les instituts de statistique et leur production. Les effets de cette évolution ne se manifesteront que progressivement, mais on peut néanmoins déjà en percevoir ou en prédire certains aspects. Les systèmes statistiques nationaux seront davantage concurrencés par des acteurs extérieurs; ils devront adapter leur façon de produire des statistiques et trouver un nouvel équilibre, en mettant à profit les nouvelles possibilités offertes par les mégadonnées. Pour cela, il sera peut-être nécessaire de passer d'un paradigme reposant sur les enquêtes à une approche généralement fondée sur la modélisation et moins directement axée sur l'acquisition de données. À l'ère des mégadonnées, la présence de spécialistes des données dans le personnel des instituts de statistique ne doit plus être l'exception, mais la règle, et la culture de ces organismes doit changer en

¹⁰ Disponible à l'adresse suivante (*en anglais uniquement*) : <http://www.statistics.gov.hk/wsc/STS018-P3-S.pdf>.

¹¹ Voir <http://www.isi2015.ibge.gov.br>.

¹² Voir <http://www.cbs-events.nl/dgins2013/>.

¹³ Voir <http://www.unece.org/stats/documents/2013.09.coll.html>.

conséquence. En repérant rapidement les possibilités offertes et les problèmes posés par le traitement massif des données, ces organismes seront bien placés pour se transformer rapidement en fournisseurs d'informations provenant de ces sources nouvelles.

14. À sa deuxième réunion, qu'il a tenue les 21 et 22 octobre 2013, le Bureau de la Conférence des statisticiens européens 2013-2014, organe directeur, dans le domaine des statistiques, de la Commission économique pour l'Europe, a examiné en détail la question des mégadonnées¹⁴. À l'issue de cet examen, il a tout particulièrement recommandé que soient précisés des domaines d'action prioritaires qui devront faire l'objet d'une collaboration, à l'échelon international, entre les différents instituts de statistique, et que soit mis en place un mécanisme d'échange d'informations sur les connaissances relatives à l'exploitation des mégadonnées et à l'expérience acquise dans ce domaine. Le Bureau a également approuvé une proposition de projet¹⁵ sur le rôle des mégadonnées, dans le but :

a) De mettre en lumière les principales possibilités offertes par les mégadonnées, de proposer des orientations aux organismes de statistique et de mettre au point une action concertée afin d'apporter une solution aux principaux problèmes stratégiques et méthodologiques que pose à la profession les mégadonnées;

b) De démontrer qu'il est possible de fournir en même temps et avec efficacité, à partir de gros volumes de données, de nouveaux produits et des statistiques officielles « classiques », et d'adapter ce modèle à différents pays;

c) De faciliter l'échange des connaissances, des compétences, des outils et des méthodes entre les différents organismes, afin de produire des statistiques à partir de mégadonnées.

15. Le thème du traitement massif des données et de la modernisation des systèmes statistiques figurait également à l'ordre du jour de quelques manifestations récemment organisées en Asie. La Division de la statistique et l'Institut national de statistiques chinois ont organisé à Tianjin (Chine), avec le soutien de la Commission économique et sociale pour l'Asie et le Pacifique (CESAP), un séminaire international sur la modernisation des instituts de statistique, du 24 au 26 octobre 2013, qui a vu l'accent mis sur la modernisation et au cours duquel plusieurs exposés ont été présentés sur les mégadonnées. Le onzième séminaire sur la gestion à l'intention des responsables des instituts nationaux de statistique des pays de l'Asie et du Pacifique¹⁶, organisé par l'Institut de statistique pour l'Asie et le Pacifique (ISAP) de la CESAP à Chiba (Japon), les 21 et 22 novembre 2013, a mis l'accent sur l'initiative pour une révolution en matière de données et sur le rôle que pouvaient jouer les mégadonnées dans le programme de développement pour l'après-2015.

¹⁴ Voir ECE/CES/BUR/2013/OCT/2.

¹⁵ Ce projet s'inscrit dans le cadre de l'initiative sur la modernisation de la production statistique, sous l'égide du Groupe de haut niveau sur la modernisation de la production et des services statistiques coordonné par la Commission économique pour l'Europe.

¹⁶ Voir http://www.unsiap.or.jp/training/6_Leading_MQS/6_ms11.php.

III. Mégadonnées : sources et données

16. Une équipe spéciale informelle établie par le Groupe de haut niveau sur la modernisation de la production et des services statistiques et coordonnée par la CEE a fait le point sur les mégadonnées et les problèmes liés à leur exploitation dans un rapport intitulé « Utilisation des “données massives” dans les statistiques officielles » (10 mars 2013), qui a été présenté à la soixante et unième session plénière de la Conférence des statisticiens européens, tenue du 10 au 12 juin 2013. Les conclusions qui y figuraient sont examinées ci-après. Les catégories proposées par l'équipe spéciale de la CEE dans le rapport ont été retenues pour le questionnaire utilisé dans le cadre de l'enquête mondiale menée par la Division de statistique au troisième trimestre de 2013 (voir sect. IV).

17. Plusieurs types de données sont aujourd'hui générés automatiquement et sans interruption sous forme numérique. Ces différentes sources de données peuvent présenter un intérêt dans le cadre des statistiques officielles pour mesurer certaines tendances sociales, environnementales, financières ou économiques avec plus de précision et plus rapidement. En règle générale, les sources de données massives peuvent être classées comme suit :

- Sources administratives (provenant de l'administration d'un programme, qu'il soit public ou non), par exemple les dossiers médicaux électroniques, les consultations et séjours à l'hôpital, les dossiers d'assurance, les dossiers bancaires et les banques alimentaires;
- Sources commerciales ou transactionnelles (provenant d'une transaction entre deux entités), par exemple les transactions au moyen de cartes de crédit et les transactions en ligne (y compris à partir de dispositifs mobiles);
- Données fournies par des détecteurs, par exemple l'imagerie satellite, les capteurs routiers et les capteurs climatologiques;
- Données des dispositifs de surveillance, par exemple des téléphones portables et du Système mondial de localisation (GPS);
- Données comportementales, c'est-à-dire recherches en ligne (sur un produit, un service ou tout autre type d'information) et pages vues en ligne;
- Données indiquant des opinions, par exemple commentaires affichés sur les réseaux sociaux.

18. Les données administratives sont l'une des principales sources utilisées par les systèmes statistiques nationaux pour établir les statistiques. Habituellement, ces données sont reçues des administrations publiques d'une manière très structurée, puis traitées, stockées, gérées et utilisées par les instituts de statistique. Les données administratives ne sont pas pour l'instant des sources de données massives, mais elles pourraient le devenir si leur vitesse ou leur volume augmente, par exemple lorsque les instituts de statistique commenceront d'utiliser plus largement les données administratives en les recueillant en temps réel ou tous les jours ou toutes les semaines au lieu du rythme annuel ou mensuel habituel.

19. L'utilisation de données massives dans la statistique officielle pose de nombreux problèmes, que l'on peut pour l'essentiel regrouper dans les catégories suivantes :

- Problèmes législatifs, concernant l'accès aux données et leur utilisation;
- Respect de la vie privée : obtenir la confiance du public et faire en sorte qu'il accepte la réutilisation des données ainsi que l'établissement de liens avec d'autres sources;
- Problèmes financiers : coût potentiel des données au regard de leurs avantages;
- Problèmes de gestion, concernant les politiques et les directives sur la gestion et la protection des données;
- Problèmes méthodologiques, concernant la qualité des données et la pertinence des méthodes statistiques;
- Problèmes technologiques, concernant les technologies de l'information.

IV. Résultats de l'enquête mondiale sur l'utilisation des mégadonnées

20. Sur la base des descriptions données dans le rapport de l'équipe spéciale de la CEE, la Division de statistique a mis au point un questionnaire sur l'utilisation des mégadonnées dans la statistique officielle, qui comptait trois grandes parties : sources, problèmes et domaines d'exploitation. Les questions concernant les sources et les problèmes rencontrés recoupaient tous les points susmentionnés, avec quelques précisions sur les aspects de gestion, décrits comme « questions de gestion, telles que l'adhésion à de nouvelles politiques et réglementations et le développement des ressources humaines avec l'ensemble des compétences et expertises nécessaires ». Cette évaluation mondiale a été entreprise pour faire œuvre d'information sur les priorités nationales, l'évolution de la situation et les expériences acquises en rapport à l'utilisation présente ou future des mégadonnées dans la statistique officielle.

21. Le questionnaire a été envoyé aux administrations statistiques de plus de 200 pays pendant la période allant de juillet à septembre 2013. Il a été envoyé en anglais en juillet 2013, en espagnol en août 2013 et en français en septembre 2013, et pouvait être complété soit en ligne, soit au format PDF. Au 2 novembre 2013, 107 réponses avaient été reçues. Tous les résultats sont mis à la disposition de la Commission de statistique dans le cadre d'un document d'information. On en trouvera ci-après un résumé.

A. Sources des mégadonnées

22. Le libellé exact de la question concernant les sources de mégadonnées était le suivant : « Veuillez s'il vous plaît indiquer si votre bureau ou d'autres organismes faisant partie du système statistique national de votre pays sont susceptibles d'utiliser les sources de mégadonnées suivantes dans les 12 prochains mois ». En cas de réponse affirmative, des détails étaient demandés sur la source de données considérée. Plus de 50 % des pays ont indiqué utiliser des sources administratives comme données massives, alors qu'ils n'étaient qu'environ 25 % à en faire autant pour chacune des cinq autres sources de données. Plusieurs pays ont demandé si de vastes dossiers administratifs devaient ou non être tenus pour des mégadonnées. Plus généralement, les données administratives jouent un rôle clef dans la

production de nombreux produits statistiques, mais encore faut-il déterminer si elles doivent être considérées dans le même contexte comme des mégadonnées.

23. Dans l'ensemble, de nombreux pays ont répondu par l'affirmative à la question de savoir s'ils utilisaient une ou plusieurs sources de mégadonnées. Toutefois, après analyse plus poussée, il s'avère que certaines de leurs réponses ne sont pas compatibles avec la définition des mégadonnées telle qu'elle est donnée dans le rapport de référence de l'équipe spéciale de la CEE. Pour donner plus de poids aux conclusions de l'enquête sur les mégadonnées, les réponses par l'affirmative ont été passées au crible et seules ont été retenues celles qui renvoyaient à des pratiques jugées pertinentes. Toutes les réponses données sans explication ou seulement avec des explications très limitées ou non satisfaisantes n'ont pas été jugées recevables. Alors qu'au départ elles représentaient 25 % à 50 % du total, les réponses positives comptaient après ajustement pour 20 % du total pour les sources de données administratives, environ 10 % pour les sources transactionnelles, les capteurs et les dispositifs de surveillance, et moins de 5 % pour les données comportementales et les données indiquant des opinions. Les résultats détaillés seront affichés sur le site Web de la Division de statistique, avec deux colonnes distinctes pour les pays ayant répondu par l'affirmative et ceux ayant fait état de « pratiques pertinentes ».

24. Les données émanant des services en charge des impôts et de la taxe sur la valeur ajoutée (TVA) figurent parmi les sources de données administratives les plus citées, avec les registres de l'état civil et du commerce. Les sources transactionnelles mentionnées sont notamment les données des supermarchés obtenues par balayage optique, les communications téléphoniques, les transactions par cartes de crédit et les données des péages. En ce qui concerne les capteurs ou sources satellitaires, plusieurs pays ont déclaré qu'ils mettaient à profit l'imagerie satellitaire dans le cadre de leurs enquêtes. Il a aussi été fait référence à l'imagerie satellitaire comme outil permettant d'étudier les statistiques de la production agricole. Les capteurs routiers sont exploités à la fois pour les statistiques du transport (indice de la circulation aux Pays-Bas) et les statistiques sur l'environnement (mesure de la qualité de l'air au Royaume-Uni de Grande-Bretagne et d'Irlande du Nord).

25. Quelques pays utilisent les sources de surveillance et les sources GPS. L'Estonie a par exemple mis à profit les données sur les téléphones mobiles pour améliorer les statistiques du tourisme, comme expliqué plus en détail dans la section suivante. Au Canada, les données GPS relatives aux itinéraires des camions et aux transports ferroviaire, maritime et aérien sont utilisées pour compléter les statistiques du transport. Les données comportementales ne sont pas exploitées par beaucoup de pays, mais certains envisagent d'y avoir recours. Par exemple, en Italie, les requêtes Internet pourraient être utilisées : a) comme source d'information complémentaire pour réduire les coûts associés aux enquêtes; b) à des fins de prévision; et c) pour affiner la précision géographique des estimations (en particulier pour les indicateurs du marché du travail).

26. De même, très peu de pays s'appuient sur les données indiquant des opinions ou qui sont tirées des réseaux sociaux pour les mégadonnées. Les Pays-Bas ont indiqué que certains chercheurs s'intéressaient aux messages affichés publiquement dans le cadre des réseaux sociaux (voir étude de cas à la section IV). À Singapour, on étudie la façon dont on pourrait, grâce à cette source de mégadonnées, repérer rapidement les problèmes de sécurité alimentaire.

B. Problèmes rencontrés dans l'utilisation des mégadonnées

27. La question exacte qui était posée était la suivante : « Ces problèmes représentent-ils un obstacle majeur pour votre système statistique national? », avec comme réponses possibles « Non (pas de problèmes majeurs) », « Sans opinion (la question n'a pas été discutée) » ou « Oui (il s'agit d'un problème majeur) ». Pour les six problèmes considérés « Oui (il s'agit d'un problème majeur) » était la réponse la plus fréquente, suivie de « Sans opinion » (ce qui révèle que ces questions n'avaient pas été suffisamment étudiées) et d'un petit nombre de « Non » (ce n'est pas un problème). Dans l'ensemble, les problèmes les plus fréquemment cités avaient trait aux méthodologies, à l'information et à la gestion, suivis de questions d'ordre législatif et de confidentialité.

28. Les répondants sont convenus qu'un des premiers enjeux était de régler les problèmes d'ordre méthodologique et qu'il faudrait promouvoir la recherche dans ce domaine. La qualité des données, leur représentativité, leur volatilité et leur dimensionnalité sont autant de difficultés méthodologiques qui compromettent l'exploitation des sources de mégadonnées. Les problèmes informatiques constituent aussi un défi majeur dans de nombreux pays parce que les systèmes de traitement et d'analyse existants ne sont pas adaptés aux mégadonnées. La nature des mégadonnées (notamment leur volume, mais aussi leur vitesse et leur variété) rend la mise au point des outils et de l'infrastructure informatiques nécessaires extrêmement délicate, en particulier dans le monde en développement. La plupart des pays ont aussi vu un défi majeur dans les questions de gestion telles que le respect des nouvelles politiques et réglementations et le développement de ressources humaines disposant des compétences et de l'expertise requise.

29. L'utilisation systématique de sources de données non gouvernementales représente un enjeu législatif pour les gouvernements dans le contexte des mégadonnées. Si certains pays ont déjà adopté des lois qui garantissent l'accès aux sources gouvernementales et non gouvernementales, d'autres ont vu dans l'absence d'une telle législation un véritable problème. Les questions de confidentialité liées à l'utilisation des mégadonnées constituent aussi un important sujet de préoccupation dans de nombreux pays. La protection des données à caractère personnel dans ce domaine est très importante et il faut donc rechercher un parfait équilibre entre la protection des données personnelles et leur utilisation statistique.

30. Plusieurs pays se sont aussi plaints de problèmes d'ordre financier, et en particulier des dépenses qu'il leur fallait engager pour trouver des sources de mégadonnées. Par ailleurs, ils ne disposaient pas toujours des ressources financières voulues pour acquérir les matériels et logiciels qui leur permettraient de traiter les mégadonnées, d'autant que les entreprises privées semblaient être de plus en plus conscientes de la valeur des données qu'elles détenaient et en augmentaient les prix.

31. Il a aussi été jugé que l'utilisation des mégadonnées par les systèmes statistiques nationaux risquait d'entraîner d'autres problèmes de taille. Ainsi, de nombreux répondants se sont inquiétés de la manière d'intégrer les mégadonnées dans les systèmes statistiques. Comment introduire cette nouvelle source de données dans les programmes statistiques existants? La plupart des répondants sont tombés d'accord pour dire que les mégadonnées ne remplaceront pas les statistiques officielles, mais viendront plutôt les compléter, par exemple en offrant des statistiques très rapidement disponibles et des indicateurs bêta pour les statistiques

clefs. Comme les mégadonnées sont un concept relativement nouveau dans le domaine des statistiques officielles, il faut aussi savoir susciter un certain enthousiasme à leur égard de la part des statisticiens officiels.

C. Domaines d'utilisation actuels et futurs

32. Dans la troisième et dernière partie de l'enquête mondiale, les répondants ont été interrogés sur les domaines dans lesquels ils utiliseraient ou envisageaient d'utiliser les mégadonnées dans les 12 prochains mois. Les choix suivants leur étaient proposés « Statistiques démographiques et sociales », « Registres et statistiques de l'état civil », « Statistiques économiques et financières », « Statistiques des prix », « Statistiques des transports », « Statistiques de l'environnement » et « Autres domaines des statistiques officielles », avec « oui » ou « non » comme réponses possibles, toute réponse par l'affirmative devant être explicitée. Les répondants ont indiqué surtout utiliser les mégadonnées pour les « statistiques démographiques et sociales » et les « statistiques économiques et financières ». Toutefois, comme c'était le cas pour l'utilisation des sources de mégadonnées, toutes les réponses affirmatives dans cette partie de l'enquête ont dû être réévaluées pour identifier les « pratiques pertinentes ». Au terme de cette réévaluation, il s'est avéré que les pratiques pertinentes en ce qui concerne l'utilisation des mégadonnées renvoyaient aux statistiques des prix et aux statistiques économiques et financières. En particulier, plusieurs pays ont cité l'utilisation de données obtenues par lecture optique ou par la technique de la récupération sur le Web pour établir des indices de prix fréquemment mis à jour qui complètent l'indice des prix à la consommation standard. Dans l'ensemble, le pourcentage de pratiques pertinentes était d'un peu plus de 10 % pour les statistiques des prix, les statistiques économiques et financières et les statistiques démographiques et sociales, et d'environ 5 % pour chacun des autres domaines d'utilisation.

33. Dans le domaine des statistiques démographiques et sociales, certains ont mentionné la façon dont on pouvait mettre à profit les données relatives aux téléphones mobiles pour examiner les comportements des voyageurs et le tourisme, les réseaux sociaux pour mesurer la confiance des consommateurs, et les données sur les vacances de poste en ligne pour étudier l'emploi, ou encore les possibilités offertes par Google. D'autres ont évoqué la coopération avec les opérateurs de téléphonie mobile pour établir des statistiques géoréférencées sur les déplacements. Dans le domaine des statistiques économiques et financières, il a été fait état de la recherche sur l'utilisation de l'imagerie satellitaire et des données des capteurs au sol pour évaluer les rendements des cultures, ainsi que de l'utilisation des données de suivi des appareils mobiles pour alimenter les statistiques sur la mobilité de la population. Un certain nombre de pays ont expliqué comment ils utilisent les mégadonnées dans le domaine des statistiques du transport. Les données détaillées provenant des péages routiers et des capteurs routiers sont aussi utilisées pour les statistiques de la circulation routière.

V. Les mégadonnées et la modernisation des systèmes statistiques

34. Il faudra moderniser les systèmes statistiques pour venir à bout des problèmes mentionnés dans les réponses à l'enquête mondiale. Il a été suggéré de mettre encore plus l'accent sur la recherche pour surmonter les difficultés méthodologiques liées à l'exploitation des sources de mégadonnées. En raison de la nature de ces mégadonnées, les questions de représentativité et de couverture de la population peuvent s'avérer problématiques. En outre, la variété et la question de la durée des données (par exemple, pendant combien de temps les gens vont continuer à utiliser Facebook) posent problème du point de vue des statistiques officielles puisque la comparabilité, la continuité et la cohérence des statistiques sont tenues pour les principaux gages de qualité. Les changements de méthodologie peuvent se traduire par un recours plus fréquent aux modélisations et nécessiter une coopération plus étroite avec le milieu universitaire. D'autres études et expérimentations doivent être menées pour explorer les applications potentielles des mégadonnées dans les statistiques officielles, la recherche risquant de devenir un élément à part entière du processus de production statistique.

35. L'infrastructure et les capacités informatiques existantes doivent être renforcées pour permettre le traitement d'ensembles de données volumineux et à grande vitesse tirés de diverses sources. Il faudrait aussi envisager d'harmoniser les méthodes et outils employés par les instituts nationaux de statistique. Il est nécessaire d'acquérir les moyens les plus pointus (logiciels, matériel et ressources humaines) pour se tenir au courant des besoins technologiques aux fins de la collecte, du traitement et de l'utilisation des mégadonnées. Les instituts doivent envisager de travailler à distance (cloud computing) plutôt que d'essayer de déplacer de grandes quantités de données sur leurs propres serveurs. Le propriétaire de la source pourrait aussi être prié de procéder aux premiers calculs et de fournir des données plus agrégées et mieux structurées à l'institut de statistique. Il faudra disposer de ressources suffisantes pour édifier l'infrastructure informatique nécessaire afin de trouver fréquemment les données pertinentes, de les lier entre elles, de les analyser et de les présenter sous un format adapté aux usagers.

36. La constitution et la rétention d'effectifs dotés des compétences requises dans les domaines de l'analyse des données et de l'informatique seront un enjeu majeur. Face aux nouvelles sources de mégadonnées, il faut revoir l'organisation de la collecte et du traitement des données. Les membres du personnel de la plupart des instituts nationaux de statistique ne sont pas suffisamment qualifiés à cet égard. Il faudrait donc renforcer les capacités des personnels des instituts nationaux et leur offrir une formation intensive pour qu'ils soient en mesure d'exploiter ces sources de données. L'expertise actuelle, fondée sur un savoir-faire statistique classique, ne garantit pas un traitement adéquat de ce nouveau type de données, et il est donc nécessaire de prévoir le recrutement et la formation de nouveaux membres du personnel ayant le profil voulu (par exemple, des scientifiques spécialisés dans les données).

37. L'adoption de nouvelles dispositions juridiques s'impose pour permettre la production de statistiques officielles à l'aide des mégadonnées. Dans de nombreux pays, le cadre législatif actuel ne couvre pas l'accès aux mégadonnées et leur utilisation, qu'elles soient d'origine gouvernementales ou proviennent du secteur privé. Il sera donc particulièrement difficile d'avoir accès aux mégadonnées

collectées et stockées par d'autres parties. En outre, un cadre de confidentialité doit être mis en place pour fixer des règles de base sur la manière dont les mégadonnées peuvent être combinées, protégées, partagées, exposées, analysées et retenues. Cela permettrait de régler la question cruciale de la confiance du public dans l'utilisation des données personnelles par les autorités gouvernementales. Il faut savoir garder la confiance des populations : les individus doivent pouvoir être certains que leurs renseignements personnels seront bien protégés et ne seront pas divulgués ou utilisés à mauvais escient, par exemple pour ce qui est des données de localisation des téléphones mobiles, une des sources potentielles de mégadonnées. Même si leur identité n'est pas divulguée, les clients des opérateurs de télécommunication mobile resteront extrêmement préoccupés par le transfert et le partage des informations les concernant dans n'importe quel but. De même, les fournisseurs d'appareils mobiles doivent recevoir des garanties de la part des gouvernements auxquels ils communiquent des données que le droit à la vie privée de leurs clients ne sera pas violé.

38. La réaction du public face à l'utilisation des mégadonnées dans les statistiques officielles sera déterminante. Les gouvernements, et à plus forte raison le secteur privé, doivent gagner la confiance du public en ce qui concerne l'utilisation des données à caractère personnel et ne pas la trahir. Le secteur public devra consentir des efforts considérables pour mettre en place un cadre d'autorisation qui définisse les règles d'accès et d'utilisation des données numériques à caractère personnel et délimite dans la transparence les rôles des dépositaires de données et des bureaux gouvernementaux compétents. C'est seulement dans un tel environnement que la vie privée et la confidentialité des renseignements personnels pourront être garanties et que le public pourra avoir confiance dans l'utilisation des mégadonnées.

VI. Études de cas

A. Australie : utilisation des images satellite dans les statistiques agricoles

39. Les images satellite sont utilisées pour étudier la couverture terrestre, ainsi que dans le cadre d'un programme de recherche qui a pour objectif de calculer le rendement de certains types de cultures à partir des données relatives à l'utilisation des terres. Les données à caractère géographique sont exploitées dans le cadre des modèles statistiques concernant l'agriculture. Elles pourraient également l'être pour les statistiques concernant la pêche, la sylviculture et l'environnement. L'accès aux données brutes et leur traitement se fait sur le serveur en nuage du système informatique du fournisseur de données, et non sur le réseau informatique de l'institut de statistique. Les algorithmes d'analyse d'image qui permettent d'extraire les données géographiques relatives à l'utilisation des terres sont conçus en collaboration avec des universitaires dans le cadre d'un projet de recherche commun concernant non seulement la mise au point d'algorithmes, mais aussi les modèles de série chronologique et géographique, les systèmes de géocodage et les procédés d'évaluation statistique.

B. Brésil : utilisation des données de surveillance à haute fréquence pour tenir les comptes de l'eau

40. En vertu d'un accord signé en 2012 entre l'Institut brésilien de géographie et de statistique (l'institut national de statistique), l'Agence nationale de l'eau (l'administration fédérale chargée de la gestion des ressources hydriques) et le Secrétariat des ressources en eau et de l'environnement urbain du Ministère de l'environnement (en charge de la coordination de la politique de ressources hydriques), un nouveau comité a été chargé de tenir les comptes de l'eau au Brésil. L'Agence nationale de l'eau surveille la quantité et la qualité des ressources hydriques grâce au réseau hydrométéorologique national, composé de près de 22 000 stations de surveillance qui recueillent des données à haute fréquence sur le débit des eaux, les précipitations, la qualité de l'eau, l'évaporation et les sédiments. L'Agence recueille ainsi chaque jour un grand nombre de données sur l'eau, qu'elle traite et diffuse gratuitement sur son site Web¹⁷ ou à partir du Système d'information national sur les ressources en eau¹⁸, dispositif complet de collecte, de traitement, de stockage et de diffusion de données et d'informations sur la gestion des ressources en eau. Pour tenir les comptes de l'eau du Brésil, l'Institut brésilien de géographie et de statistique complète les données recueillies dans le cadre de son activité normale de collecte de statistiques économiques et environnementales avec celles fournies par le Système d'information.

C. Bhoutan : utilisation des outils de téléphonie mobile pour établir l'indice des prix à la consommation

41. Au Bhoutan, la collecte des données sur les prix à la consommation est appuyée et facilitée par les outils modernes de communication. L'indice des prix à la consommation est en effet établi chaque mois à partir des données recueillies dans les divers districts par les statisticiens locaux au moyen d'une application Android pour téléphone portable. Dans chaque district, les fonctionnaires se rendent dans les magasins pour y collecter des données sur tous les produits composant le panier de l'indice et les transmettent automatiquement au serveur situé au siège du Bureau national de statistique à Thimphou, qui combine toutes les données provenant des différents districts pour calculer l'indice des prix et le taux d'inflation au niveau national.

42. La quantité de données traitées au Bhoutan au moyen de cette application n'est pas très importante pour l'instant, mais elle pourrait le devenir si l'utilisation des appareils portables et de la transmission sans fil dans la collecte des prix se généralisait. La méthode suivie au Bhoutan, qui consiste à commencer par recueillir une quantité limitée de données électroniques, pourrait en outre servir de modèle pour d'autres pays en développement. L'entrée automatisée des données chiffrées et leur transfert sur un serveur central au moyen de technologies sans fil constituent une première étape indispensable avant de pouvoir exploiter les sources de mégadonnées.

¹⁷ <http://www2.ana.gov.br>.

¹⁸ <http://www.snirh.gov.br/telemetria/consultardadoshidrovisualizacaogeo/inicial.action>.

D. Colombie : utilisation des images satellite dans les statistiques agricoles et les statistiques des transports

43. La Colombie utilise les images satellite comme une source de mégadonnées dans plusieurs domaines. Par exemple, dans le cadre d'un essai pilote, les images satellite sont mises à profit pour compléter les résultats des enquêtes de recensement agricole. L'essai, dans sa phase initiale, vise à déterminer le type de variables qui peuvent être extraites des images satellite et venir confirmer les informations recueillies lors des recensements. Par ailleurs, l'Office des Nations Unies contre la drogue et le crime et le Ministère de l'intérieur et de la justice, dans le cadre du système intégré de surveillance des cultures illicites, utilisent les images satellite pour surveiller les champs de coca et évaluer leur production.

44. L'Institut national des routes de Colombie utilise les données du système GPS pour améliorer la circulation routière et affiner les statistiques des transports. Grâce à un nouveau système, les voitures n'ont pas à s'arrêter aux postes de péage; en lieu et place, un dispositif électronique de localisation installé à bord du véhicule est détecté à l'entrée du péage. Ce type de péage assure un haut niveau de confort, de sécurité et de service. Le dispositif de localisation contient en outre toutes les informations concernant le véhicule, qui complètent celles figurant au registre national unique des transports. Ce nouveau système a été à ce jour mis à l'essai dans 10 gares de péage et il a déjà permis d'améliorer le contrôle de la circulation et les statistiques des transports.

E. Estonie : utilisation des données de localisation des téléphones portables dans les statistiques des voyages internationaux

45. La méthode d'établissement des statistiques des voyages internationaux, qui utilise la localisation passive par téléphone portable, a été mise au point conjointement par la Banque centrale d'Estonie, l'Université de Tartu et Positium LBS, une société qui recueille des données anonymes auprès des opérateurs de téléphonie mobile, les traite au moyen de modèles statistiques et généralise à l'ensemble des données les résultats obtenus à partir d'échantillons. Cette méthode fournit des informations d'ensemble fiables sur les personnes qui franchissent la frontière estonienne pour entrer dans le pays ou le quitter. Les données provenant de la localisation par téléphone mobile sont anonymes. Elles sont traitées, analysées et présentées d'une façon qui ne permet pas d'identifier les numéros de téléphone ou leurs titulaires.

46. La Banque centrale utilise les statistiques des déplacements internationaux pour calculer la valeur des importations et des exportations de services de voyages dans le compte courant de la balance des paiements. Dans le respect de la méthodologie de la balance des paiements, la localisation par téléphone portable permet de déterminer le lieu de résidence des voyageurs, indépendamment de leur nationalité, grâce au critère du domicile permanent. Le lieu de résidence probable du détenteur du téléphone est établi par son utilisation des services d'itinérance et des réseaux de téléphonie locaux. Le nombre de voyages effectués par les détenteurs de téléphone (résidents et non-résidents), leur durée et leur nature sont établis par l'utilisation qu'ils font de leur téléphone, selon le pays et la région où ils se trouvent.

47. Les statistiques des voyages à l'étranger prennent en compte les déplacements d'une journée ou moins effectués à l'étranger par les personnes qui résident en Estonie. Il se peut qu'une personne se rende dans plusieurs pays lors d'un même voyage. Est considéré comme pays de destination le pays le plus éloigné de l'Estonie, ou le (ou les) pays dans lequel (ou lesquels) le voyageur passe au moins deux jours consécutifs. Les pays dans lesquels le voyageur ne passe pas la nuit ou qui ne répondent pas aux critères du pays de destination sont considérés comme des pays de transit. Le nombre de voyages effectués par les détenteurs de téléphone (résidents et non-résidents) et leur durée sont calculés au moyen de modèles statistiques, puis les résultats obtenus sont généralisés à l'ensemble des données, en prenant en compte également les autres sources de données officielles (nombre de personnes transitant par le port et l'aéroport de Tallin, nombre de personnes franchissant la frontière orientale de l'Estonie, statistiques officielles de l'hébergement).

48. Les données anonymes provenant de la localisation par téléphone portable sont utilisées conformément aux lois estoniennes et aux textes de l'Union européenne sur la protection des données, notamment des données à caractère personnel (loi estonienne sur la protection des données à caractère personnel; directive 95/46/CE du Parlement européen et du Conseil du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données; directive 2002/58/CE du Parlement européen et du Conseil du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques). Afin de garantir la protection des données à caractère personnel, c'est en coopération avec le bureau d'inspection pour la protection des données que la méthode de localisation par téléphone portable a été développée et utilisée dans divers domaines en Estonie depuis 2006.

F. Pays-Bas : les médias sociaux comme source de données potentielle pour les statistiques officielles

49. De plus en plus de personnes utilisent les médias sociaux pour échanger des informations, discuter de questions qui les intéressent et correspondre avec leur famille et leurs amis. Aux Pays-Bas, les médias sociaux sont utilisés par une part importante de la population : environ 70 % des habitants ont indiqué s'exprimer sur au moins une des plateformes. Le taux de réponse aux enquêtes de l'institut néerlandais de statistique ne cessant de baisser, l'idée est venue d'utiliser les messages publiés et diffusés volontairement sur les médias sociaux comme une source de données pour les statistiques officielles. On a ainsi étudié les messages en libre accès publiés sur diverses plateformes, comme Twitter ou Facebook, ainsi que les messages publics affichés sur les sites d'information, les forums et les blogues. Ces messages ont été obtenus auprès d'une société qui collecte systématiquement tous les messages en libre accès écrits en néerlandais sur la partie néerlandophone du Web.

50. Ce sont à la fois la teneur et le ton des messages qui ont été étudiés. Les études sur le contenu des messages en néerlandais sur Twitter, la plateforme de média social où sont publiés le plus de messages en néerlandais en libre accès, ont montré que près de 50 % des messages constituaient des « bavardages sans intérêt ». Les autres portaient essentiellement sur les activités de loisir (10 %), le travail (7 %), les

médias (télévision et radio) (5 %) et la politique (3 %). Mais l'utilisation des messages plus instructifs était compliquée par l'existence des remarques frivoles sans intérêt. Il est apparu que l'analyse du sentiment général qui ressort de l'ensemble des messages publiés sur tous les médias sociaux en libre accès pouvait présenter un fort intérêt statistique puisque l'on a constaté une étroite corrélation entre la gestion générale et la confiance des consommateurs néerlandais, et notamment le jugement qu'ils portaient sur la situation économique. Cette corrélation, bien que très variable à l'échelle de la journée, était stable à l'échelle de la semaine ou du mois, d'où l'idée d'établir des indices d'humeur hebdomadaires et mensuels comparables à l'indice de confiance des consommateurs. Ces indices peuvent être publiés le premier jour ouvrable suivant la semaine étudiée et prouver qu'il est possible d'obtenir rapidement des résultats.

VII. Conclusions et perspectives

51. Les statisticiens ont pris conscience des possibilités qu'offrent les mégadonnées dans le domaine des statistiques officielles. Elles peuvent les aider à mieux accomplir leur tâche, qui est de fournir des statistiques à jour et cohérentes sur l'économie, la société et l'environnement afin d'améliorer la prise de décisions et de nourrir la recherche et le débat public. Par ailleurs, grâce au développement des techniques de géocodage des données économiques, sociales et environnementales, et du fait qu'un nombre croissant d'informations peuvent être extraites de clichés géographiques à très basse résolution, des progrès rapides sont attendus dans l'utilisation des mégadonnées dans les pays développés et en développement. Il convient donc, dans le cadre d'initiatives régionales, que les instituts de statistique des différents pays mettent en commun leurs innovations méthodologiques, leurs meilleures pratiques en rapport aux questions stratégiques et leurs moyens de formation, afin de pouvoir exploiter au mieux les mégadonnées et de résoudre notamment les problèmes qui se posent en matière de cadre législatif, de confidentialité, de coûts, de gestion, de méthodes et de techniques. Tout en traitant ces questions, les États et les entreprises devront mettre en place un cadre qui préserve le caractère privé et confidentiel des informations personnelles utilisées dans les statistiques officielles, de façon à ce que l'exploitation des mégadonnées dans les statistiques officielles puisse gagner la confiance du public.

52. L'évaluation mondiale sur l'utilisation des mégadonnées dans les statistiques officielles décrit quelques exemples intéressants, notamment des cas où des statistiques ont été établies à partir de données recueillies lors de transactions ou au moyen d'appareils de localisation, de capteurs ou de satellites. Ainsi, des statistiques des prix ont été établies à partir d'informations recueillies par lecture optique; des statistiques des dépenses de consommation, à partir de transactions électroniques par carte; des statistiques de voyages et des transports, à partir d'appareils de localisation par GPS; des statistiques sur le bien-être, à partir de données recueillies sur les médias sociaux; et des statistiques agricoles, à partir d'images satellite. L'étude mondiale témoigne également des nombreux problèmes rencontrés dans cette activité, notamment les difficultés méthodologiques dans la compilation des mégadonnées, les importants moyens informatiques nécessaires et les problèmes de confidentialité.

53. Pour répondre aux défis que posent les mégadonnées, le système statistique devra de toute évidence être modernisé. Il faudra en particulier développer les activités de recherche, renforcer les moyens informatiques, réorganiser les équipes pour y intégrer des spécialistes de données et nouer des partenariats avec le secteur privé pour la mise en commun des données numériques générées automatiquement, dans un cadre respectueux de la vie privée des personnes et des règles de confidentialité.

54. Les défis que posent les mégadonnées et la modernisation des systèmes statistiques, ainsi que les possibilités qu'elles offrent, sont semblables dans la plupart des pays. Les instituts nationaux de statistique pourraient donc utilement mettre en commun leurs expériences, leurs pratiques et leurs solutions. Quelques initiatives en ce sens sont déjà engagées au niveau régional. Une proposition de projet de la CEE visant à renforcer le rôle des mégadonnées dans la modernisation de la production de statistiques a été approuvée dans le cadre de la Conférence des statisticiens européens. Des initiatives semblables sont en cours dans la région Asie-Pacifique sous l'impulsion de la CESAP, et le lancement de projets aux ambitions similaires est envisagé dans d'autres régions.

55. Il est envisagé de créer un groupe de travail mondial sur les mégadonnées, dont les activités viendraient soutenir les ambitions et les initiatives régionales dans le cadre d'un programme de travail mondial, qui permettrait la mise en commun des pratiques entre commissions régionales. Tandis que celles-ci auraient vocation à travailler sur des projets particuliers avec un petit nombre de pays au niveau régional, le groupe de travail mondial gèrerait les problèmes rencontrés et les ferait connaître pour favoriser la conclusion d'un accord sur la méthodologie, les meilleures pratiques en rapport aux questions stratégiques, la formation et l'utilisation des plateformes communes. Dans le cadre de sa mission, le groupe de travail aurait à faire valoir que les systèmes statistiques nationaux, qui ont pour tâche de valider l'utilisation des mégadonnées à des fins pratiques et d'intégrer les mégadonnées dans les statistiques officielles existantes, sont plus à même de tenir ce rôle que les autres parties.

56. Le groupe de travail pourrait en outre s'intéresser plus particulièrement aux questions concernant la place des mégadonnées dans l'élaboration du programme de développement pour l'après-2015 et aurait pour tâche de favoriser la coopération avec les entités des Nations Unies, les organisations internationales, les organisations de la société civile et les autres membres du partenariat mondial en faveur de la révolution des données. L'une des critiques faites aux indicateurs des objectifs du Millénaire pour le développement est qu'ils ne sont pas à jour. Le traitement des mégadonnées se faisant le plus souvent en temps réel, il pourrait être possible de publier des indicateurs beaucoup plus récents.

57. Une conférence internationale se tiendra au deuxième ou au troisième trimestre de 2014 pour lancer le groupe de travail mondial et définir ses activités concernant la méthodologie commune, la formation, les plateformes et les autres questions relatives à l'utilisation des mégadonnées dans les statistiques officielles.

VIII. Questions soumises à la Commission pour examen

58. La Commission de statistique est invitée à donner son avis sur les points suivants :

a) S'il y a lieu d'étudier plus avant les sources de mégadonnées, les problèmes qu'elles posent et les domaines dans lesquels elles sont utilisées pour établir des statistiques officielles à l'échelon mondial, notamment au regard de la situation des pays en développement et en relation avec le programme de développement pour l'après-2015 et l'initiative en faveur d'une révolution des données;

b) La création à l'échelon mondial d'un groupe de travail sur l'utilisation des mégadonnées dans les statistiques officielles, dont les activités viendraient compléter celles des commissions régionales et qui s'occuperait de questions d'importance mondiale.
