



Conseil économique et social

Distr. générale
12 décembre 2014
Français
Original : anglais

Commission de statistique

Quarante-sixième session

3-6 mars 2015

Point 3 a) iii) de l'ordre du jour provisoire*

Questions soumises à la Commission

pour examen et décision : données

à l'appui du programme de développement

pour l'après-2015 : mégadonnées

Rapport du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle

Note du Secrétaire général

En application de la décision 2014/219 du Conseil économique et social, le Secrétaire général a l'honneur de transmettre le rapport du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle. Ce rapport présente les éléments saillants de la Conférence internationale sur l'utilisation des mégadonnées en statistique officielle, les conclusions de la première réunion du Groupe et les résultats de l'enquête menée sur l'utilisation des mégadonnées en statistique officielle. Il décrit les attributions et le projet de programme de travail du Groupe portant sur la formation, les compétences et le renforcement des capacités, l'articulation des mégadonnées et des objectifs de développement durable, la sensibilisation et la communication, l'accès aux données et les partenariats, mais aussi sur des questions transversales, telles que les classifications et les cadres de référence. Il étudie également certaines sources de mégadonnées utilisées en statistique officielle, à savoir les données issues de la téléphonie mobile et des médias sociaux et l'imagerie satellitaire. Il s'achève par l'énumération des points soumis à examen.

La Commission est invitée à prendre acte du rapport.

* E/CN.3/2015/1.



Rapport du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle

I. Introduction

1. Du fait de l'utilisation constante et généralisée des télécommunications et des autres systèmes et appareils issus de l'innovation technologique, tels que les dispositifs de géolocalisation par satellite (GPS), les guichets automatiques de banque, les numériseurs, les capteurs, les téléphones portables, les satellites et les médias sociaux, des données numériques sont produites en permanence. Couramment appelées « mégadonnées », ces informations, qui peuvent être utiles en statistique officielle, présentent un volume, une vitesse de circulation et une diversité si importants que de nouveaux outils et de nouvelles méthodes sont nécessaires pour pouvoir en assurer avec efficacité le recueil, la gestion et le traitement. Or, de nombreux bureaux de statistique n'ont pas encore monté de dossier argumentaire démontrant les avantages qu'il y aurait à investir dans cette nouveauté et exposant pourquoi et dans quelles conditions les mégadonnées présentent une utilité et un intérêt. La Conférence internationale sur l'utilisation des mégadonnées en statistique officielle, évoquée plus amplement ci-dessous, s'est avérée encourageante à cet égard, en offrant de multiples exemples favorables de projets menés actuellement en la matière.

2. Les possibilités que recèlent les mégadonnées tiennent à ce qu'elles donnent accès très rapidement, voire en temps réel, à de grandes quantités de données, qui sont habituellement produites pour un coût modique. Les méthodes traditionnelles d'acquisition des données que constituent, par exemple, les enquêtes menées auprès des ménages ou auprès des entreprises, demandent du temps et représentent souvent un poste important de dépenses dans la production des données. L'utilisation des mégadonnées pourrait permettre de compléter ou de réduire ces opérations de recueil de données, voire s'y substituer. Toutefois, les statisticiens n'ignorent pas qu'avant de commencer à utiliser les mégadonnées dans les statistiques officielles, et pour pouvoir tirer parti de ces nouvelles méthodes d'acquisition, et notamment les employer pour suivre la réalisation des objectifs de développement durable et rendre compte des progrès réalisés en la matière, il leur faudra d'abord s'attacher à régler de façon satisfaisante les questions qu'elles posent sur les plans de la méthodologie, de la qualité, de la technologie, de l'accès aux données, de la législation, de l'intimité de la vie privée, de la gestion et des finances, et présenter des analyses coûts-avantages appropriées.

3. C'est ainsi que, lors de sa quarante-cinquième session, qui a eu lieu du 4 au 7 mars 2014, la Commission de statistique a rappelé que les mégadonnées constituent une source d'information dont on ne peut faire abstraction et qui doit être évaluée objectivement¹. Dans cette optique, elle a souscrit à la proposition tendant à mettre en place un groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle, chargé de dresser l'inventaire des activités en cours et des exemples d'utilisation des mégadonnées, de répondre aux préoccupations relatives à la méthodologie, aux ressources humaines, à la qualité et à la confidentialité, et d'élaborer des directives pour classer les différents types de sources de mégadonnées.

¹ Voir E/2014/24, chap. I, sect. B, décision 45/110.

4. La Commission a demandé instamment que les attributions et le mandat du Groupe soient fondés sur des considérations stratégiques et s'articulent en particulier avec le programme de développement pour l'après-2015, l'initiative sur la révolution des données et les Principes fondamentaux de la statistique officielle. Elle a estimé que le Groupe avait également vocation à compléter et à mettre à profit les travaux effectués par les commissions régionales et d'autres organismes statistiques internationaux, et que, dans son programme de travail, il devait accorder une attention particulière à la situation des pays en développement, et notamment à l'état de leur cadre juridique et aux lacunes que peuvent présenter leurs équipements informatiques.

5. Le Groupe a été créé en mai 2014, conformément à la décision 45/110 de la Commission. Il s'est réuni pour la première fois le 31 octobre 2014 à Beijing, dans la foulée de la Conférence internationale sur l'utilisation des mégadonnées en statistique officielle qui y avait été coorganisée par la Division de statistique du Département des affaires économiques et sociales de l'ONU et le Bureau national des statistiques chinois. On notera, par ailleurs, que la Division et la Commission économique pour l'Europe s'étaient associées pour réaliser une enquête sur les projets relatifs à l'utilisation des mégadonnées et les structures dans lesquels ils s'inscrivent, qui a servi de matière à ces rencontres et a été utile à leur préparation.

6. Le chapitre ci-après offre une vue d'ensemble des travaux de la Conférence internationale. Il est suivi d'un compte rendu succinct de la première réunion du Groupe (chap. III), puis de la présentation des résultats de l'enquête réalisée sur les projets relatifs à l'utilisation des mégadonnées et le cadre dans lequel ils s'insèrent (chap. IV). Enfin, le rapport s'achève par l'énumération des points sur lesquels la Commission est invitée à se prononcer.

II. Conférence internationale sur l'utilisation des mégadonnées en statistique officielle

7. Quelque 120 participants de plus de 40 pays différents, issus de diverses organisations régionales et internationales, mais aussi du monde de la recherche, des milieux universitaires et du secteur privé, ont participé à la Conférence internationale sur l'utilisation des mégadonnées en statistique officielle. Leurs travaux ont porté essentiellement sur trois catégories de sources de mégadonnées, à savoir : a) les téléphones portables, les systèmes de navigation GPS et autres dispositifs de localisation; b) les images satellitaires et autres données géoréférencées; et c) Twitter et les autres médias sociaux. Les participants ont dressé la typologie de chacune de ces sources, examiné leurs enjeux particuliers et évoqué les partenariats nécessaires à leur exploitation. Les séances de clôture ont été l'occasion d'aborder les avantages et inconvénients communs aux sources de mégadonnées, les moyens à mettre en œuvre pour faire adopter l'innovation qu'elles représentent et la marche à suivre pour établir un dossier argumentaire en faveur de leur utilisation.

8. L'importance du nombre des projets présentés lors de la Conférence témoigne de l'esprit novateur qui anime les acteurs de la statistique. Les applications tirées de l'utilisation de diverses sources de mégadonnées dans un large éventail de domaines statistiques prouvent bien qu'elles sont à même d'améliorer les statistiques officielles. Ces outils innovants peuvent venir enrichir l'arsenal des systèmes statistiques en complétant les applications en vigueur ou en offrant des solutions rapides et plus souples à des enjeux très actuels. Au vu des différents projets menés,

on comprend mieux également comment aborder certaines difficultés touchant à la méthodologie, à l'accès, à l'intimité de la vie privée et aux compétences. Les paragraphes ci-après reviennent sur des temps forts de la Conférence.

9. Les caractéristiques techniques de l'enregistrement des données de téléphonie mobile, et notamment les différences existant entre les données de localisation active et passive des téléphones portables, ont été abordées. Plusieurs applications statistiques et expériences nationales intéressantes, ayant pour objet d'établir des statistiques du tourisme et des déplacements diurnes, de disposer d'évaluations en matière de recensement démographique, de dresser des cartes de la pauvreté ou de suivre l'évolution des déplacements en cas d'épidémies, ont été présentées. Compte tenu de la généralisation de l'utilisation des appareils mobiles, y compris dans le monde en développement, les données des téléphones portables ont toutes les chances d'apporter des informations peu coûteuses en temps réel sur des questions de grand intérêt relatives au développement. Les principales difficultés à résoudre en la matière sont la protection de la confidentialité, l'accès aux données des téléphones portables et la confiance du public dans leur utilisation.

10. Les images satellites peuvent être un excellent moyen d'obtenir, plus souvent et plus rapidement, des données présentant un degré de détail très poussé, notamment dans le domaine des statistiques agricoles. Toutefois, les méthodes de mesure employées, notamment pour estimer les rendements agricoles, sont encore en phase d'essai. Les travaux menés actuellement par le Bureau australien de statistique sur les données issues de l'imagerie satellitaire pourraient permettre, à terme, de compléter et de remplacer en partie les enquêtes utilisées à l'heure actuelle pour mesurer la production agricole. Grâce à l'utilisation des images satellitaires produites tous les 15 jours, il serait ainsi possible de diminuer la fréquence des enquêtes et de réduire les frais y afférents. Des applications analogues de l'imagerie satellitaire en statistique officielle sont actuellement à l'étude en Chine, en Colombie et au Mexique ainsi que dans d'autres pays, où, par exemple, la comptabilité des écosystèmes fait l'objet de tests et d'expérimentations.

11. Twitter, Facebook et les autres médias sociaux étant, sans doute, détenteurs des plus grands ensembles de données existants sur les comportements humains, les statisticiens les étudient dès à présent dans l'espoir d'aboutir, grâce à leur exploitation, à des applications en sciences sociales, par exemple en matière de santé. Les Pays-Bas offrent à cet égard un exemple très prometteur de mesure des perceptions des consommateurs à partir des données de Facebook et de Twitter, permettant d'obtenir des estimations préliminaires et d'avoir moins recours aux enquêtes, les estimations tirées des médias sociaux pouvant être produites plus souvent et à moindre coût. En Chine et en Italie, des outils d'aspiration de données (*web scraping*) sont utilisés pour évaluer les taux de vacance de postes. Ils peuvent servir à étayer les statistiques du travail existantes en permettant d'améliorer les prévisions mensuelles et d'affiner les estimations territoriales. Les statisticiens réalisent de grands progrès dans ce domaine, même s'il leur reste encore à résoudre de nombreux problèmes de méthode.

12. Les similitudes des différentes sources de données ont également été évoquées, notamment sur les plans de la méthode et de la qualité, mais aussi en ce qui concerne la confidentialité, les partenariats et les technologies de l'information. Il est apparu que toutes les sources de mégadonnées offrent, d'une manière ou d'une autre, des difficultés qu'il est, avant toute chose, indispensable de connaître avec

précision. Une fois la problématique des sources de mégadonnées mieux cernée, les statisticiens pourront alors tirer parti de leurs points communs. Ainsi, des modèles d'accords-cadres appelés à être conclus avec des entreprises exerçant leurs activités au niveau mondial pourraient être mis au point pour l'accès aux données, indépendamment de la source concernée. Les statisticiens devraient non pas chercher à rivaliser avec le secteur privé, mais collaborer avec lui, afin de tirer tout le parti possible des mégadonnées en statistique officielle. Il y a lieu cependant qu'ils conservent leur impartialité et leur indépendance, en s'employant à faire connaître les avantages que présente l'exploitation des ressources des données numériques disponibles au profit de la société. La confiance du public sera, en effet, indispensable pour réussir.

13. Enfin, les débats ont également porté sur les missions actuelles et futures des acteurs de la statistique publique. Ainsi, les statisticiens officiels pourraient être plus souvent appelés à valider les informations fournies par les acteurs du marché. Il importe, cela a été souligné, que les acteurs de la statistique restent pénétrés de la mission qui leur incombe de fournir des informations de qualité et impartiales sur des questions intéressant la société dans son ensemble et les décideurs en particulier. Si l'innovation joue un rôle déterminant et si, à ce titre, il appartient aux statisticiens de s'en emparer, les activités traditionnelles qui font le cœur de leur métier demeurent, et resteront indispensables. Ainsi, il faudra évaluer les résultats des applications tirées de l'utilisation des mégadonnées et, à cet effet, il sera nécessaire de recourir à des enquêtes traditionnelles pour dégager les points de référence nécessaires.

14. Il convient que les statisticiens mettent au point des dossiers argumentaires sur l'utilisation des mégadonnées. Il leur faudra en particulier veiller à articuler l'utilisation des mégadonnées avec le programme de développement pour l'après-2015 et ses objectifs de développement durable, qui mettront à rude épreuve les systèmes statistiques au cours des années qui viennent. Si les nombreuses cibles proposées dans le cadre du programme de développement pour l'après-2015 ne pourront, bien entendu, pas toutes être mesurées à l'aide d'indicateurs fondés sur les mégadonnées, celles-ci ont cependant vocation à jouer un rôle important, compte tenu de la rapidité de leur production et de leur degré de précision géospatiale. L'une des pistes les plus prometteuses est d'étudier la mise au point d'indicateurs supplétifs, faisant appel aux mégadonnées, qui pourraient permettre d'obtenir des informations plus fréquemment que les enquêtes. Ainsi, les mégadonnées viendraient compléter les systèmes traditionnels, sans s'y substituer, de façon analogue aux estimations instantanées du produit intérieur brut (PIB), par exemple.

15. Pour résumer, la Conférence a été l'occasion de dresser un panorama des applications les plus prometteuses des mégadonnées en statistique officielle et de mieux comprendre les difficultés qu'il faudra aplanir en matière de méthode, de confidentialité et d'accès. Elle a permis de favoriser les échanges d'expériences entre les bureaux de statistique nationaux et d'autres parties prenantes et acteurs importants sur les problèmes touchant à la mise en place des partenariats, à l'accès aux données, à la communication et à la sensibilisation, comme aux besoins en matière de formation, de perfectionnement et de renforcement des capacités.

III. Réunion du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle

16. À sa première réunion tenue le 31 octobre 2014, le Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle a examiné son mandat (voir annexe I) et son programme de travail pour 2015. Le Groupe de travail mondial comprend actuellement 28 membres représentant des pays développés et des pays en développement et divers organismes internationaux et régionaux (voir annexe II).

17. En marge de la conférence, la Division de statistique, la CEE, la Commission économique et sociale pour l'Asie et le Pacifique (CESAP), l'Institut de statistique pour l'Asie et le Pacifique, l'Union internationale des télécommunications (UIT) et EUROSTAT se sont réunis pour coordonner les travaux des organismes internationaux sur les mégadonnées. Tous les organismes se sont engagés à partager leurs informations et leurs résultats dans ce domaine. La CEE établira et tiendra à cet effet un calendrier des réunions. Il est entendu qu'un des moyens de stimuler la coopération sera d'instaurer des liens étroits entre les sites Web consacrés par les différents organismes aux mégadonnées.

A. Mandat du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle

18. D'une manière générale, la communauté statistique est tenue d'étudier l'utilisation de nouvelles sources de données afin de répondre aux attentes de la société quant à de meilleurs produits et à des méthodes de travail plus efficaces. Toutefois, le mandat du Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle devrait aussi tenir compte des nouvelles demandes de données découlant de l'obligation de suivi et de communication de l'information imposée par le programme de développement pour l'après-2015.

19. L'utilisation des mégadonnées en statistique officielle est pleinement approuvée dans les Principes fondamentaux de la statistique officielle. Ainsi, les principes 1, 5 et 6 énoncent respectivement ce qui suit : a) des organismes responsables de la statistique officielle doivent établir les statistiques officielles selon un critère d'utilisation pratique et les rendre disponibles, en toute impartialité, en vue de rendre effectif le droit d'accès des citoyens à l'information publique; b) les données utilisées à des fins statistiques peuvent être tirées de toutes sortes de sources, qu'il s'agisse d'enquêtes statistiques ou de fichiers administratifs, et les organismes responsables de la statistique doivent choisir leur source en tenant compte de la qualité des données qu'elle peut fournir, de leur actualité, des coûts et de la charge qui pèse sur les répondants; c) les données individuelles recueillies pour l'établissement des statistiques par les organismes qui en ont la responsabilité, qu'elles concernent des personnes physiques ou des personnes morales, doivent être strictement confidentielles et ne doivent être utilisées qu'à des fins statistiques. Selon ces principes, les organismes de statistique ont non seulement la possibilité mais aussi pour ainsi dire l'obligation d'enquêter sur l'utilisation des sources de mégadonnées à des fins statistiques et d'utiliser eux-mêmes ces sources, du moment

que les droits au respect de la vie privée et à la confidentialité des informations personnelles sont strictement respectés. Les mégadonnées présentent l'avantage d'être pertinentes et d'actualité, et leurs moyens de recueil sont plus économiques que les méthodes de collecte des données classiques.

20. Le mandat du Groupe de travail mondial découle aussi des conclusions du rapport du Groupe consultatif d'experts indépendants sur la révolution des données pour le développement durable², intitulé « A world that counts: mobilising the data revolution for sustainable development » (novembre 2014). Dans ce rapport, il est souligné que les bureaux de statistique devront évoluer, et plus rapidement que par le passé, et continuer de s'adapter, renoncer aux méthodes de production coûteuses et contraignantes, adopter de nouvelles sources de données, y compris les données administratives d'autres institutions publiques, et s'employer à fournir des données qui soient lisibles par l'homme et par la machine et compatibles avec les systèmes d'information géospatiale et à les rendre disponibles assez rapidement de manière à faire correspondre leur cycle de vie avec le cycle décisionnel³. Dans son rapport, le Groupe consultatif d'experts indépendants fait aussi expressément référence aux mégadonnées et affirme que les capacités nationales en matière de science des données doivent être renforcées afin d'exploiter toutes les possibilités que recèlent les mégadonnées et de compléter les statistiques officielles. Un renforcement des moyens au niveau national et l'appui international aux pays en développement sont nécessaires pour mettre la révolution des données au service du développement durable. Il faut mettre au point, grâce à la coordination et à la collaboration bilatérales et internationales, des applications de données communes et standardisées pour la statistique officielle et les développer en toute transparence et en pleine conformité avec les lois en vigueur⁴.

21. Dans son rapport, le Groupe consultatif a proposé, entre autres choses, la création d'un ensemble de réseaux destinés à promouvoir l'innovation des données, l'objectif étant de rassembler les organismes et les experts dans ce domaine. Cette démarche contribuerait à l'adoption de pratiques optimales en vue d'un meilleur suivi des nouveaux objectifs de développement durable, permettrait de cerner les domaines où des infrastructures de données communes pourraient remédier au manque de moyens et renforcer l'efficacité, encouragerait la collaboration, permettrait de recenser les lacunes importantes en matière de recherche et stimulerait l'innovation.

22. Pour résumer, les organismes de statistique devraient choisir les sources de données en fonction de leur qualité, de leur actualité, des coûts et de la charge qui pèsent sur les répondants. Les sources de mégadonnées satisfont aux critères souhaités à cet égard car, comme il est indiqué plus haut, les mégadonnées ont l'avantage d'être pertinentes et d'actualité, leur collecte est moins coûteuse que la collecte des données classiques et il est possible de faire correspondre le cycle de vie des données au cycle décisionnel. Les travaux menés sur les mégadonnées devraient contribuer à l'adoption de pratiques optimales et à l'amélioration du suivi

² Le Secrétaire général a constitué le Groupe consultatif d'experts indépendants le 29 août 2014 et l'a chargé de rédiger un rapport sur la révolution des données dans le contexte du programme de développement pour l'après-2015. Le Groupe consultatif a publié son rapport le 6 novembre 2014.

³ « A World that counts », p. 9.

⁴ « A World that counts », p. 23.

des nouveaux objectifs de développement durable du programme de développement pour l'après-2015. Certains des nouveaux indicateurs ou indicateurs supplétifs pourraient être basés sur des sources de mégadonnées, ce qui les rendrait plus actuels et permettrait une ventilation sociale et géographique.

23. C'est dans ce contexte que le Groupe de travail mondial a défini ses attributions et son mandat, qui consisteraient à : a) assurer l'orientation stratégique, la direction et la coordination d'un programme mondial sur l'utilisation des mégadonnées en statistique officielle, notamment pour la mesure des indicateurs relatifs au programme de développement pour l'après-2015; b) promouvoir l'utilisation concrète des sources de mégadonnées, y compris les données transfrontalières, en s'appuyant sur des précédents et en trouvant des solutions aux nombreuses difficultés qui se posent; c) favoriser le renforcement des capacités, la formation et la mise en commun des données d'expérience; d) encourager la communication et la sensibilisation en matière d'utilisation des mégadonnées à des fins concrètes, en particulier pour le suivi du programme de développement pour l'après-2015; e) raffermir la confiance du public dans l'utilisation des mégadonnées du secteur privé en statistique officielle.

24. Le mandat du Groupe de travail mondial est présenté dans son intégralité dans l'annexe I.

B. Programme de travail et réalisations escomptées

25. Comme suite aux débats tenus durant la Conférence internationale et à sa propre réunion, le Groupe de travail mondial a décidé que son programme de travail couvrirait un certain nombre de volets d'activités qui seraient supervisés par des équipes, sous la coordination générale d'une équipe supplémentaire. Il a été décidé en premier lieu de poursuivre les travaux sur les trois catégories de sources de mégadonnées mises en avant à la Conférence internationale, à savoir les données provenant de la téléphonie mobile, l'imagerie satellite et les données issues des médias sociaux. Il était entendu que pour progresser dans leurs travaux, les trois équipes chargées de ces volets d'activités devraient exécuter au moins un projet pilote.

26. Deuxièmement, sachant que l'un des thèmes récurrents de la Conférence internationale était l'accès aux données et l'établissement de partenariats avec le secteur privé et les autres communautés, une équipe chargée d'étudier en profondeur ces deux questions connexes a été créée. Cette équipe étudiera aussi la possibilité d'établir des accords généraux sur l'accès aux données avec les fournisseurs de mégadonnées qui opèrent à l'échelle mondiale. L'accès aux mégadonnées et leur utilisation sont aussi liés à la confiance du public et à la communication des avantages et des inconvénients que présentent les mégadonnées en général. Par conséquent, il a été créé une équipe chargée de la communication et de la sensibilisation, qui examinera aussi des stratégies de collecte de fonds visant à susciter la participation active des pays en développement aux projets pilotes.

27. Troisièmement, dans la perspective du programme de développement pour l'après-2015, il a été décidé de confier à une équipe la tâche spéciale de surveiller l'établissement de liens entre les indicateurs relatifs aux objectifs de développement durable et les applications des mégadonnées. Pour finir, il a été proposé et décidé de créer deux autres équipes : l'une chargée de la formation, du perfectionnement des

compétences et du renforcement des capacités (sachant que le traitement des mégadonnées requiert des compétences autres que celles utilisées actuellement dans les bureaux de statistique); et la deuxième chargée des questions transversales comme le classement et les orientations. Cette dernière s'occupera notamment du classement des sources de mégadonnées en s'inspirant du classement proposé par le groupe de travail sur les mégadonnées de la CEE. Les travaux de ces huit équipes seraient coordonnés par une équipe composée des chefs des huit équipes.

28. Il a été indiqué que ces équipes devraient s'inspirer de l'expérience acquise dans le cadre du projet sur les mégadonnées de la CEE et de la documentation produite par les équipes de travail sur les mégadonnées de la CEE, puisque celles-ci sont sur le point d'achever leur travail alors que les équipes du Groupe de travail mondial commencent seulement le leur. Cela leur permettrait de continuer les travaux sur leur lancée. Chaque équipe du Groupe de travail mondial devra établir son propre programme de travail, se fixer des délais et définir des résultats à atteindre, en concertation étroite avec l'équipe de coordination.

29. Comme les réalisations escomptées du Groupe de travail mondial sont directement liées aux travaux de ses huit équipes, le Groupe de travail mondial devra en principe formuler un certain nombre de propositions et produire des rapports, comme indiqué en détail ci-après, et participer également à divers projets pilotes. Il continuera aussi de dresser l'inventaire des projets relatifs à l'utilisation des mégadonnées, probablement par l'intermédiaire de son équipe de coordination. Les résultats de la première enquête sont présentés dans la section suivante. Les réalisations concrètes attendues en 2015 sont les suivantes :

- Des propositions et activités concernant la formation, le perfectionnement des compétences et le renforcement des capacités aux fins de l'utilisation des mégadonnées en statistique officielle;
- Un rapport sur les liens entre les mégadonnées et les objectifs de développement durable;
- Une stratégie de communication et de sensibilisation en matière d'utilisation des mégadonnées en statistique officielle;
- Un rapport sur l'utilisation des données provenant de la téléphonie mobile en statistique officielle, qui donne des explications et des solutions sur les problèmes méthodologiques et informatiques et ceux liés au respect de la vie privée, notamment sur l'utilisation éventuelle de ces données pour le suivi du programme de développement pour l'après-2015;
- Un rapport sur l'utilisation des données issues des médias sociaux en statistique officielle, qui donne des explications et des solutions sur les problèmes méthodologiques et informatiques et ceux liés au respect de la vie privée, notamment sur l'utilisation éventuelle de ces données pour le suivi du programme de développement pour l'après-2015;
- Un rapport sur l'utilisation des images satellites et des données de télédétection en statistique officielle, qui donne des explications et des solutions sur les problèmes méthodologiques et informatiques et ceux liés au respect de la vie privée, notamment sur l'utilisation éventuelle de ces données pour le suivi du programme de développement pour l'après-2015;

- Des propositions sur le renforcement de l'accès aux sources de mégadonnées et des partenariats avec le secteur privé;
- Un rapport sur des questions transversales, le classement, les orientations et la taxonomie.

30. Outre les réalisations concrètes de ses différentes équipes, le Groupe de travail mondial participera à un certain nombre de projets pilotes sur l'utilisation des mégadonnées en statistique officielle et continuera d'établir un inventaire des projets relatifs à l'utilisation des mégadonnées, en se fondant sur l'inventaire de la CEE et sur les résultats de la première enquête effectuée par la Division de statistique et la CEE, qui sont présentés dans la section suivante.

IV. Résultats de l'enquête réalisée par la Division de statistique et la CEE sur les stratégies et l'inventaire des projets concernant l'utilisation des mégadonnées

31. En septembre 2014, la Division de statistique et la CEE ont réalisé une enquête conjointe sur les projets relatifs à l'utilisation des mégadonnées en statistique officielle. Il s'agissait de dresser une vue d'ensemble des projets en cours, afin que la communauté dans son ensemble puisse en discuter en meilleure connaissance de cause, et de poursuivre l'élaboration des programmes du Groupe de travail mondial. Les supports d'enquête ont été envoyés à tous les bureaux participant au projet de la CEE dans ce domaine et aux membres du Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle.

32. L'enquête avait pour double objectif de recueillir des informations sur les projets achevés, en cours ou potentiels en matière d'utilisation des mégadonnées, notamment au sujet des partenariats, des sources de données et des outils, et de rassembler des éléments pour mieux comprendre la place occupée par ces projets au sein des organismes concernés. Deux questionnaires distincts ont été distribués. Le premier visait à obtenir des renseignements généraux sur l'organisation, les stratégies d'utilisation et les structures de gouvernance des mégadonnées, tandis que le second portait sur les projets à proprement parler. Aux fins de cette enquête, une définition relativement large des mégadonnées a été adoptée qui consiste à les décrire comme des sources produisant à un rythme soutenu de grands volumes de données très variées dont la capture, la conservation, la gestion et le traitement nécessitent de nouveaux outils et procédés.

33. En tout, 78 bureaux de statistique nationaux et 28 organisations internationales⁵ ont été sollicités. Trente-deux pays ont répondu au questionnaire traitant du contexte d'utilisation, tandis que 24 ont complété le questionnaire de projet individuel, ce qui a permis d'obtenir des informations sur un total de 54 projets. Pour leur part, les organisations internationales ont communiqué des renseignements au sujet de trois projets. Il semble raisonnable de considérer que ces pays et ces organisations internationales représentent probablement un nombre conséquent des projets relatifs à l'utilisation des mégadonnées qui ont été achevés

⁵ Les documents ont été communiqués aux organisations internationales par courrier électronique, essentiellement pour qu'elles soient informées de l'enquête.

ou sont cours dans le monde. Néanmoins, le Groupe de travail mondial envisage d'effectuer une enquête de suivi dans l'optique de recueillir des informations concernant d'autres projets aux fins de l'inventaire des projets relatifs à l'utilisation des mégadonnées. Le rapport analytique complet sur les conclusions de cette enquête est fourni en tant que document de base. On trouvera ci-après une synthèse où sont reprises certaines des principales conclusions.

A. Principales conclusions : contexte d'utilisation

34. Le questionnaire traitant du contexte d'utilisation était axé sur l'expérience générale et les cadres de traitement des mégadonnées, en termes de stratégie, de structures de gouvernance et de gestion, d'évaluation de la qualité, de confidentialité et de pénurie de compétences. Les réponses reçues corroborent les conclusions de la Conférence internationale sur l'utilisation des mégadonnées en statistique officielle quant aux grands défis qui se posent aux bureaux de statistique nationaux.

35. Il ressort de l'enquête que si seuls quelques pays se sont dotés d'une vision à long terme de l'utilisation des mégadonnées, ils sont en revanche un certain nombre à s'apprêter à formuler une stratégie en la matière. De surcroît, les États ne sont pas rares qui ont mis en place des laboratoires, des équipes spéciales ou des groupes de travail internes pour exécuter des projets pilotes destinés à déterminer si et comment les mégadonnées pourraient être employées comme source de statistiques officielles, mais la plupart n'ont pas encore défini de processus métier pour intégrer dans leurs activités les sources de mégadonnées et les résultats produits et ne disposent pas non plus de structure propre à la gestion des projets relatifs à l'utilisation des mégadonnées.

36. Des mesures sont actuellement prises dans bien des pays pour établir des partenariats aux fins d'étudier, dans le cadre de la formulation d'une stratégie pertinente, les possibilités offertes par les mégadonnées. En vue d'atténuer au mieux les risques d'échec associés à l'exploration de ces nouvelles technologies, certains pays ont décidé de prendre une part active à des initiatives de collaboration internationale comme le projet Sandbox (« bac à sable ») de la CEE, l'équipe spéciale d'EUROSTAT sur les mégadonnées et le Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle, créé en application d'une décision de la Commission de statistique.

37. Rares sont les pays qui ont adopté un cadre spécifique aux mégadonnées en ce qui concerne le respect de la vie privée; dans leur grande majorité, les répondants ont indiqué qu'à l'heure actuelle les dispositions prévues en la matière pour les statistiques classiques s'appliquaient également aux mégadonnées. Compte tenu de l'importance que revêt leur image publique, nombre d'organisations soulignent combien il est crucial, dans le traitement des mégadonnées, de veiller au respect de la vie privée et de garantir la confidentialité, au-delà même de ce qui est strictement prévu par la loi. Certaines organisations évitent la question de la vie privée et de la confidentialité en effectuant tous les essais et toutes les manipulations dans les locaux du fournisseur de mégadonnées pour ne transmettre que les agrégats au bureau de statistique national. Il y en a par ailleurs qui estiment que les règles strictes de protection des données, bien que nécessaires, compliquent souvent l'accès aux données.

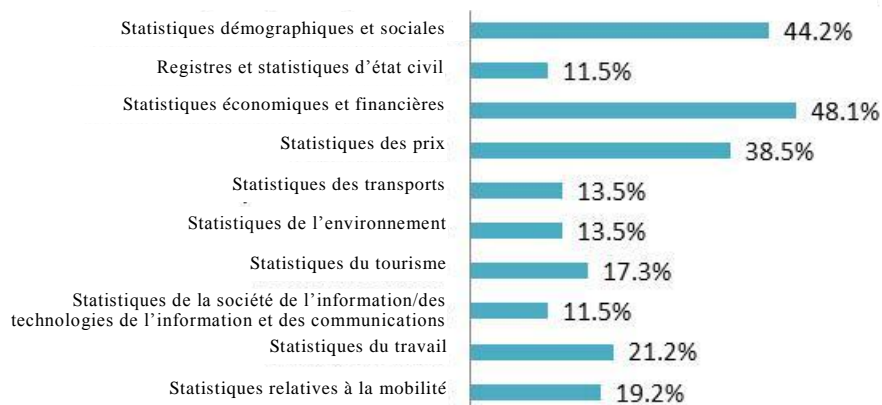
38. Tandis que la plupart des répondants reconnaissent les difficultés liées à l'informatique, aux compétences, à la législation et à la méthodologie, le plus grand défi rencontré par la majorité des projets relatifs à l'utilisation des mégadonnées tient à l'accès limité ou restreint aux ensembles de données potentiels. Du fait que les mégadonnées sont, dans une large mesure, détenues par le secteur privé (par exemple : sociétés en lignes, opérateurs de téléphonie mobile et banques), il devient très important d'entretenir avec lui une collaboration étroite. Nombre de ces acteurs sont des entreprises mondiales; la communauté statistique mondiale pourrait donc mettre à profit son pouvoir de négociation pour s'assurer l'accès à ces ensembles de données.

B. Principales conclusions : projets individuels

39. Le questionnaire consacré aux projets individuels comportait des questions plus détaillées sur les domaines d'utilisation potentiels, l'état d'avancement des travaux, les résultats et les enseignements tirés, ainsi que sur les accords de partenariat, les sources de données, l'analyse, les compétences et les outils employés. Beaucoup des projets pour lesquels des informations ont été recueillies dans le cadre de l'enquête étaient en cours d'exécution et menés afin de mettre à l'essai et d'expérimenter l'utilisation des mégadonnées. Quelques-uns avaient pris fin. Les autres projets se trouvaient à un stade exploratoire plus ou moins avancé.

40. En termes de domaines d'utilisation potentiels, la plupart des projets englobent des statistiques économiques et financières, des statistiques démographiques et sociales et des statistiques des prix (voir la figure ci-dessous). Il convient de noter que nombre de projets sont applicables à plusieurs domaines statistiques. De tels projets exploitent un type unique de sources de données qu'ils utilisent dans différents domaines statistiques. Par exemple, l'Institut national de statistique et de géographie du Mexique a recueilli un échantillon de messages affichés sur Twitter et s'est ensuite servi des données ainsi collectées pour produire différents types de statistiques, à savoir des indicateurs de bien-être subjectif et des statistiques de tourisme et de mobilité aux frontières.

Domaines d'utilisation potentiels pour ce projet



41. Un tout petit peu plus de la moitié des projets reposent sur des partenariats, le partenaire le plus courant étant une entreprise commerciale, suivie par d'autres organismes publics. Le partenariat est le plus souvent établi avec un fournisseur de données dans le but de s'assurer l'accès aux données. Cette catégorie inclut les partenariats avec des sociétés de téléphonie mobile, des fournisseurs de données recourant aux compteurs intelligents et des agrégateurs de données issues des médias sociaux. Par exemple, dans le cadre d'un projet consacré à la production de schémas de mobilité quotidienne, l'Institut national de statistique italien s'est associé à un fournisseur de données de téléphonie mobile. En Suisse, l'Office fédéral de la statistique a noué un partenariat avec les quatre grandes enseignes de commerce de détail du pays pour obtenir deux fois par mois des données provenant des scanners de caisse afin de produire des indices des prix à la consommation en ce qui concerne les denrées alimentaires et apparentées.

42. En plus de s'associer avec des fournisseurs de données commerciales, les bureaux de statistique nationaux établissent également des partenariats avec des prestataires de services d'analyse pour de nombreux projets. Cela pourrait avoir pour but d'acquérir le savoir-faire spécifique de la source de mégadonnées, de transformer les données brutes pour les inscrire dans la structure voulue ou d'appliquer des techniques analytiques de manière à produire certains résultats. Toutefois, ce genre de partenariat pourrait aussi être conclu dans l'optique de disposer de ressources humaines capables d'exécuter des tâches qui demandent beaucoup de travail. Par exemple, l'Institut national de statistique italien a mis en place une collaboration avec une université mexicaine afin de s'adjoindre l'aide d'un grand nombre d'étudiants pour classer par contenu les publications affichées sur le réseau Twitter, en vue d'analyser le bien-être subjectif.

43. Dans leur vaste majorité, les participants à l'enquête ont décidé d'opter pour des solutions d'hébergement internes plutôt que de recourir à des services proposés sur le marché. Ils ont cité des préoccupations relatives au respect de la vie privée comme raison principale de ce choix. Les hôtes externes sont plus souvent utilisés lorsque les données sont réputées appartenir déjà au domaine public. Néanmoins, l'hébergement interne pourrait devenir un facteur limitatif pour le choix d'une source de mégadonnées, dans la mesure où les investissements requis devraient couvrir le matériel, les logiciels et les compétences humaines, ce qui serait susceptible de constituer un obstacle de taille au lancement d'un projet relatif à l'utilisation des mégadonnées, en particulier du fait de l'évolution rapide de la technologie.

V. Conclusions et questions à examiner

44. La communauté statistique a commencé à étudier sérieusement les possibilités offertes par les sources de mégadonnées dans le domaine de la statistique officielle. La Conférence internationale sur l'utilisation des mégadonnées en statistique officielle qui s'est tenue à Beijing a mis en avant les applications de plusieurs sources de mégadonnées pour différents types de statistiques officielles et favorisé le partage d'expérience entre pays développés et pays en développement. Nombre des difficultés rencontrées par les bureaux de statistique nationaux sont similaires – tels que l'accès aux données, les préoccupations relatives au respect de la vie privée, les problèmes d'ordre méthodologique, les besoins accrus en termes de capacités de calcul et le manque de compétences en interne pour l'analyse des

mégadonnées – et soulignent la nécessité d’une collaboration mondiale. Le Groupe de travail mondial a débuté ses travaux par la formulation de ses modalités de fonctionnement et des domaines qu’il se propose d’aborder en priorité. Il a également entamé la préparation d’un inventaire des activités en cours qui ont trait à l’utilisation des mégadonnées en statistique officielle. L’enquête sur les projets d’exploitation des mégadonnées a permis de recenser des exemples intéressants dont les riches enseignements pourraient bénéficier à d’autres organisations. Il est essentiel que le Groupe de travail mondial, en collaboration avec les groupes de travail régionaux, élabore des orientations complémentaires et mène un certain nombre de projets pilotes en vue de consolider et d’accélérer les progrès en matière d’utilisation des mégadonnées en statistique officielle.

45. La Commission de statistique est invitée à faire connaître ses vues sur :

a) Les attributions et le mandat du Groupe de travail mondial sur l’utilisation des mégadonnées en statistique officielle;

b) Les équipes spéciales qu’il est proposé de créer dans le cadre du Groupe de travail mondial et les réalisations correspondantes en ce qui concerne la formation, les compétences et le renforcement des capacités; l’établissement d’un lien entre les mégadonnées et les objectifs de développement durable; les activités de plaidoyer et la communication; les données issues de la téléphonie mobile; les données recueillies par l’intermédiaire des médias sociaux; l’imagerie satellitaire; l’accès et les partenariats; et les questions intersectorielles, telles que les classifications et les cadres de travail;

c) Les résultats de l’enquête réalisée par la Division de statistique de la Commission économique pour l’Europe et les plans visant à la transposer à l’échelle mondiale;

d) Les projets pilotes que le Groupe de travail mondial entend mener en collaboration avec les groupes de travail régionaux afin de consolider et d’accélérer les progrès en matière d’utilisation des mégadonnées en statistique officielle.

Annexe I

Attributions et mandat du Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle

Il est fait à la communauté statistique l'obligation d'étudier l'utilisation possible des nouvelles sources de données pour répondre aux attentes de la société en termes d'amélioration des produits et de renforcement de l'efficacité des méthodes de travail. Les attributions et le mandat du Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle reposent sur ce principe, mais devraient certainement aussi être compris dans le sens de la satisfaction de la demande de nouvelles données qui découlera des prescriptions de suivi et de communication prévues au titre du programme de développement pour l'après-2015.

La lacune principale des indicateurs de suivi des objectifs du Millénaire pour le développement tenait à un problème de délais (ou de disponibilité). Pour éviter cet écueil avec les objectifs définis dans le cadre du programme de développement pour l'après-2015, le Groupe de personnalités de haut niveau chargé d'étudier le programme de développement pour l'après-2015 a appelé, dans son rapport du 30 mai 2013 (« Pour un nouveau partenariat mondial : vers l'éradication de la pauvreté et la transformation des économies par le biais du développement durable »), à une révolution des données, qui puiserait dans les sources de données, nouvelles et existantes, pour intégrer pleinement les statistiques dans la prise de décisions, promouvoir un libre accès aux données et encourager leur utilisation, et mobiliser un appui accru au profit des systèmes statistiques.

Le présent mandat fait également référence au rapport du Groupe consultatif d'experts indépendants sur la révolution des données pour le développement durable, où il est souligné que les bureaux de statistique devront évoluer et continuer de s'adapter, notamment en renonçant aux processus de production lourds et coûteux, en incorporant de nouvelles sources de données, y compris des données administratives provenant d'autres services gouvernementaux, et en se concentrant sur la mise à disposition de données qui soient lisibles par les utilisateurs et les machines, compatibles avec les systèmes d'information géospatiale et disponibles assez rapidement pour garantir que le cycle de données concorde avec le cycle de décision^a.

Dans ce contexte, les sources de mégadonnées sont reconnues comme un élément important de la révolution des données qui s'impose pour assurer le suivi des objectifs de développement pour l'après-2015. L'utilisation des mégadonnées pourrait contribuer à améliorer certains aspects de la qualité des statistiques, tels que l'opportunité et la pertinence, sans que cela soit au détriment de l'impartialité ni de la rigueur méthodologique.

En outre, les Principes fondamentaux de la statistique^b officielle encouragent l'utilisation de nouvelles sources de données telles que les mégadonnées. Il y est en effet indiqué ce qui suit :

^a « A World that counts: mobilising the Data Revolution for sustainable development », p. 9.

^b Voir résolution 68/261 de l'Assemblée générale.

- Des organismes responsables de la statistique officielle doivent établir les statistiques officielles selon un critère d'utilité pratique et les rendre disponibles, en toute impartialité, en vue de rendre effectif le droit d'accès des citoyens à l'information publique (principe 1);
- Les données utilisées à des fins statistiques peuvent être tirées de toutes sortes de sources, qu'il s'agisse d'enquêtes statistiques ou de fichiers administratifs. Les organismes responsables de la statistique doivent choisir leur source en tenant compte de la qualité des données qu'elle peut fournir, de leur actualité, des coûts et de la charge qui pèse sur les personnes sondées (principe 5);
- Les données individuelles recueillies pour l'établissement des statistiques par les organismes qui en ont la responsabilité, qu'elles concernent des personnes physiques ou des personnes morales, doivent être strictement confidentielles et ne doivent être utilisées qu'à des fins statistiques (principe 6).

Sur la base de ces considérations stratégiques énoncées dans les Principes fondamentaux, du programme de développement pour l'après-2015 et de la révolution des données, et compte tenu de la décision 45/110 de la Commission de statistique, le mandat du Groupe de travail mondial est formulé comme suit :

a) Élaborer une vision stratégique, définir des orientations et mener des activités de coordination aux fins d'un programme mondial sur l'utilisation des mégadonnées en statistique officielle, notamment pour les indicateurs du programme de développement pour l'après-2015;

b) Promouvoir une utilisation pratique des sources de mégadonnées, y compris les données transfrontières, tout en mettant à profit les précédents existants et en trouvant des solutions aux nombreux problèmes qui se posent, parmi lesquels :

- Les problèmes d'ordre méthodologique, qui ont trait à la qualité et aux données;
- Les problèmes juridiques et autres en ce qui concerne l'accès aux sources de données;
- Les problèmes relatifs au respect de la vie privée, en particulier ceux qui se rapportent à l'utilisation et à la réutilisation des données, à leur mise en correspondance et à leur réidentification;
- Les problèmes liés à la sécurité et aux technologies de l'information, y compris les moyens sophistiqués de diffusion des données, l'évaluation de l'informatique et du stockage en nuage et l'analyse coût-avantage;

c) Promouvoir également le renforcement des capacités, la formation et le partage d'expérience;

d) Encourager la communication et les activités de plaidoyer en faveur de l'utilisation des mégadonnées à des fins stratégiques, en particulier pour le suivi du programme de développement pour l'après-2015;

e) Améliorer la confiance du public dans l'utilisation des mégadonnées en statistique officielle.

Annexe II

Membres du Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle

Pays

Australie
Bangladesh
Cameroun
Chine
Colombie
Danemark
Égypte
Émirats arabes unis
États-Unis d'Amérique
Indonésie
Italie
Maroc
Mexique
Pays-Bas
Oman
Pakistan
Philippines
République-Unie de Tanzanie

Organisations

Banque mondiale
Centre de statistique du Conseil de coopération des États arabes du Golfe
Commission économique et sociale pour l'Asie et le Pacifique/Institut de statistique pour l'Asie et le Pacifique
Commission économique pour l'Europe
Division de statistique du Département des affaires économiques et sociales
EUROSTAT
Initiative Global Pulse
Organisation de coopération et de développement économiques
Union internationale des télécommunications
Union postale universelle