



Conseil économique et social

Distr. générale
19 décembre 2018
Français
Original : anglais

Commission de statistique

Cinquantième session

5-8 mars 2019

Point 4 e) de l'ordre du jour provisoire*

Questions soumises pour information : utilisation des mégadonnées en statistique officielle

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

Note du Secrétaire général

Conformément à la décision 2018/227 du Conseil économique et social et à la pratique établie, le Secrétaire général a l'honneur de transmettre le rapport du Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle. Dans le présent rapport, le Groupe de travail mondial donne suite aux demandes que la Commission de statistique a formulées à sa dernière session, notamment celle de mettre des produits et services à la disposition des statisticiens du monde entier sur la plateforme mondiale de données, méthodes et outils d'apprentissage fiables pour la statistique officielle, en répondant aux inquiétudes concernant la protection de la vie privée et la confidentialité des données et en précisant davantage le modèle d'activité de la plateforme mondiale. Celle-ci est un environnement de recherche-développement qui a pour vocation de mettre en commun et à l'essai des méthodes, des données et des matériels pédagogiques fiables. Elle offre une infrastructure technologique et des services en statistique officielle auxquels le secteur privé, les universitaires et la société civile peuvent collaborer. En outre, les différentes équipes spéciales du Groupe de travail mondial (sur l'observation de la Terre, les téléphones mobiles, les données issues des réseaux sociaux et saisies par lecture optique, et les techniques de protection de la vie privée) ont organisé des ateliers de formation, établi des manuels et collaboré à des projets novateurs. La Commission est invitée à prendre note du présent rapport.

* E/CN.3/2019/1.



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

I. Introduction

1. La Commission de statistique a créé le Groupe de travail mondial sur l'utilisation des mégadonnées en statistique officielle à sa quarante-cinquième session, en 2014. Conformément à son mandat (voir [E/CN.3/2015/4](#)) et à la décision 46/101 de la Commission (voir [E/2015/24-E/CN.3/2015/40](#)), le Groupe de travail assure 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 aux objectifs de développement durable du Programme de développement durable à l'horizon 2030.

2. Dans sa décision 49/107 (voir [E/2018/24-E/CN.3/2018/37](#)), la Commission de statistique a réaffirmé que l'utilisation des mégadonnées et d'autres sources nouvelles de données nouvelles était essentielle à la modernisation des institutions statistiques nationales, afin que celles-ci gardent leur raison d'être dans un domaine en évolution rapide, et souligné que les mégadonnées pouvaient servir à combler des lacunes, à améliorer la rentabilité des opérations statistiques, à renouveler les enquêtes et à accroître la granularité des produits. La Commission a soutenu la proposition du Groupe de travail visant à continuer de développer une plateforme mondiale participative de recherche-développement de données, de méthodes et d'outils d'apprentissage fiables, rappelé qu'il était nécessaire de présenter l'étude de viabilité de la plateforme, encouragé le Groupe de travail à faire fond sur les résultats obtenus jusque-là en mettant des produits et services de portée pratique à la disposition du système statistique mondial, afin de favoriser la production de statistiques et d'indicateurs, notamment les indicateurs relatifs aux objectifs de développement durable, et souligné qu'il était nécessaire d'examiner avec soin les défis sociétaux que représentent la confiance, la déontologie, le respect de la vie privée ainsi que la confidentialité et la sécurité des données.

3. La section II du présent rapport appelle l'attention sur la réunion annuelle du Groupe de travail mondial et sa journée portes ouvertes, qui ont eu lieu à Doubaï le 21 octobre 2018. À l'occasion de la journée portes ouvertes, des sessions ont été organisées sur l'utilisation de la plateforme mondiale de données, méthodes et outils d'apprentissage fiables pour la statistique officielle et sur les travaux des différentes équipes spéciales du Groupe de travail mondial¹. La Section III fournit quelques informations supplémentaires sur la plateforme et les résultats obtenus par les équipes spéciales, notamment sur les activités de renforcement des capacités, et la Section IV présente les mesures que le Groupe de travail doit prendre pour avancer dans son programme de travail.

II. Réunion annuelle du Groupe de travail mondial et journée portes ouvertes

4. La journée portes ouvertes sur la plateforme mondiale ayant été organisée le même jour que la cinquième réunion annuelle du Groupe de travail mondial, moins

¹ Un rapport sur la réunion peut être consulté sur le site Web du Groupe de travail mondial à l'adresse <https://unstats.un.org/bigdata/>.

de temps a été consacré à cette réunion que les années précédentes. Les points suivants ont néanmoins été examinés : l'organisation de la cinquième Conférence internationale sur l'utilisation des mégadonnées, qui devrait se tenir à Kigali au printemps 2019, les progrès réalisés dans la mise en place de la plateforme mondiale, les questions relatives à la gestion des données devant être traitées par le Groupe de travail, s'agissant notamment de la gestion de la plateforme mondiale, un bref compte rendu des travaux des équipes spéciales et la préparation du rapport du Groupe de travail à la Commission de statistique.

5. Les années précédentes, la réunion annuelle du Groupe de travail mondial durait une journée entière, juste avant l'ouverture de la Conférence internationale sur l'utilisation des mégadonnées, comme ce fut le cas à Beijing (2014), Abou Dhabi (2015), Dublin (2016) et, plus récemment, à Bogota (2017). Les comptes rendus de ces réunions, ainsi que les comptes rendus et documents relatifs aux réunions du Bureau du Groupe de travail, peuvent désormais être consultés sur le site Web du Groupe de travail². À l'origine, la cinquième réunion annuelle du Groupe de travail devait se tenir avant la cinquième Conférence internationale sur l'utilisation des mégadonnées, qui, selon le calendrier traditionnel, devait avoir lieu à l'automne 2018. Cependant, après l'annonce de la tenue du deuxième Forum mondial des Nations Unies sur les données à Doubaï (Émirats arabes unis), fin octobre 2018, le Bureau a décidé que la Conférence internationale aurait lieu au printemps, et pour commencer, à Kigali, en 2019. Toutefois, le Groupe de travail avait quand même besoin de tenir sa réunion annuelle afin d'éviter d'avoir une période de 12 mois sans réunion physique de l'ensemble de ses membres. Celle-ci a donc été exceptionnellement organisée la veille de l'ouverture du Forum mondial des Nations Unies sur les données.

6. Les deux points principaux ayant retenu l'attention du Groupe de travail mondial sont la mise en place de la plateforme mondiale et les questions correspondantes relatives à la gestion des données. La plateforme mondiale n'est plus un concept mais une réalité, qui se traduit par la fourniture de données, de méthodes et d'outils d'apprentissage. À cet égard, le Groupe de travail doit s'entendre sur les concepts de ses quatre piliers fondamentaux – données fiables, méthodes fiables, partenariats fiables et outils d'apprentissage fiables – et les définir plus précisément. Cela implique de trouver un accord sur les questions de la propriété de divers grands jeux de données et de leurs modalités d'accès sur la plateforme mondiale, ainsi que sur celles de la liberté d'accès aux données et aux algorithmes, et du degré d'indépendance par rapport à la plateforme des logiciels, services et outils mis à disposition. Ces questions ont une incidence directe sur le modèle d'activité de la plateforme. Le Groupe de travail a décidé qu'un document portant d'une part sur ce modèle et d'autre part sur la gestion des données serait élaboré en étroite collaboration, respectivement, avec l'Office for the National Statistics (Bureau de la statistique nationale) du Royaume-Uni de Grande-Bretagne et d'Irlande du Nord et Statistique Canada.

7. La journée portes ouvertes sur la plateforme mondiale a été organisée par le Groupe de travail dans les locaux de l'Autorité fédérale de la compétitivité et de la statistique des Émirats arabes unis. Elle a consisté en une démonstration du fonctionnement de la plateforme, suivie de sessions sur les statistiques de productions agricoles et de couverture du sol (à partir de données satellitaires), la mesure de la mobilité humaine (à partir de données issues de la téléphonie mobile), la mesure des

² Voir <https://unstats.un.org/bigdata/bureau/>.

fluctuations de prix (à partir de données saisies par lecture optique) et sur les techniques de protection de la vie privée³.

8. La plateforme permet à la communauté statistique mondiale de collaborer sur des projets dans le domaine des données. Elle abrite un nombre croissant de jeux de données, notamment celles issues des satellites Landsat et Sentinel, les données satellitaires expérimentales de Planet.com, les données AIS de position des navires et les données ADS-B de position des aéronefs. Elle offre des services tels que des serveurs infonuagiques, des services d'analyse géospatiale et l'application Jupyter Notebook. Le Data Science Campus de l'Office for the National Statistics a utilisé la plateforme mondiale pour mener à bien diverses études, notamment pour établir l'indice de végétation urbaine de 112 villes du Royaume-Uni (à partir de 17 millions d'images fournies par Google Street View) ou évaluer la proportion de la population rurale vivant à moins de 2 kilomètres d'une route praticable toute l'année (indicateur 9.1.1 des objectifs de développement durable) en Irlande du Nord (en utilisant Open Street Map et des données démographiques).

9. Les sessions ultérieures de la journée portes ouvertes ont été l'occasion de présenter l'utilisation de nouvelles sources de données et de nouvelles technologies en statistique officielle, en montrant comment utiliser la plateforme mondiale pour réaliser certains projets. À terme, le Groupe de travail souhaite que davantage de projets fassent un usage intensif de la plateforme en y conservant les méthodes expérimentées et les données, en vue de leur mise en commun. Le Bureau de statistique de la Colombie a évalué, dans le cadre d'un projet pilote, le rendement des cultures céréalières en appliquant un modèle déterministe qui intègre les résultats d'un traitement d'images satellitaires et d'autres sources de données. Entre autres choses, on utilise les algorithmes de Google Earth pour le prétraitement et le traitement des images satellitaires en vue d'obtenir l'indice de l'état de la végétation et l'indice de l'état des températures, sans lesquels on ne peut évaluer le rendement des cultures.

10. Statistique Canada améliore ses modèles de rendement de cultures vivrières à l'aide de données satellitaires. Des essais d'estimation de rendements ont été effectués pour 19 cultures, sur le mois de septembre, à l'aide de données satellitaires. Pour 15 de ces cultures, les résultats ont été jugés suffisamment satisfaisants pour être publiés dans les statistiques officielles. Cela signifie que les enquêtes sur le rendement des cultures pourraient être progressivement remplacées par des estimations reposant sur des données satellitaires. Finalement, Statistique Canada a conclu qu'il fallait accélérer l'apprentissage et créer les conditions permettant aux intéressés de faire des expériences et d'évaluer leur qualité. L'avantage de la plateforme mondiale, c'est qu'elle facilite la collaboration à l'aide de méthodes, de données et de partenariats fiables.

11. Des agents du Programme des Nations Unies pour l'environnement ont présenté l'application Global Surface Water Explorer qui offre un accès libre et gratuit à des données agrégées sur l'étendue d'eau par pays, bassin versant et bassin secondaire et permet de mesurer l'indicateur 6.6.1 des objectifs de développement durable. Cela peut inciter à s'engager davantage, dans les zones où d'autres sources de données sont inexistantes. En outre, l'application assure la comparabilité dans le temps et dans

³ On trouvera des détails à l'adresse <https://unstats.un.org/unsd/bigdata/conferences/2018/open-day/default.asp>, sous l'onglet « Agenda ».

l'espace et offre un aperçu sur le passé. Elle tire parti des meilleures compétences, à l'échelle mondiale, en matière d'algorithmes d'observation de la Terre.

12. Le moteur de traitement et d'analyse Positium permet au bureau de statistique de l'Indonésie d'améliorer la qualité de ses statistiques de tourisme. Grâce à des données de positionnement mobile, les lacunes des statistiques de mouvements transfrontières de touristes venant de pays voisins ont pu être comblées. L'utilisation de ces mêmes données a permis d'évaluer le nombre et la destination des voyages de tourisme intérieur avec plus de précision, et ce, à l'aide de mesures plus fréquentes et à moindre coût.

13. Eurostat est en train de mettre au point une approche méthodologique unifiée du traitement des données issues de la téléphonie mobile en statistique officielle. Un cadre méthodologique de référence, comme on l'appelle communément, facilitera les relations entre les opérateurs de réseaux mobiles et les statisticiens sur les plans technique et organisationnel, et permettra d'assurer la cohérence, la reproductibilité et la transférabilité des méthodes de traitement. Cette coopération permettra également de préciser des aspects juridiques, dans le cadre notamment du Règlement général sur la protection des données, en s'appuyant sur des éléments concrets, et d'effectuer le traitement et l'analyse de données issues d'opérateurs de réseaux mobiles multiples (autrement dit, de fusionner des données provenant de différents opérateurs).

14. Eurostat propose de travailler à trois niveaux : celui des données des opérateurs de téléphonie mobile où s'effectue le traitement des données brutes ; celui du traitement statistique où sont appliquées les méthodes statistiques ; et un niveau intermédiaire, celui de la convergence, où les données confidentielles issues des téléphones mobiles rencontrent les méthodes statistiques. À ce dernier niveau, les statisticiens peuvent concevoir des algorithmes et les opérateurs de téléphonie mobile les exécuter en recourant au calcul multipartite sécurisé.

15. Les migrations internationales sont un des sujets les plus débattus de la vie politique, ce qui se traduit par une demande sans précédent de statistiques dans ce domaine. La Division de statistique du Département des affaires économiques et sociales du Secrétariat aide les pays à améliorer leurs capacités de collecte et de diffusion de statistiques sur les flux et les stocks de migrants ; elle collabore avec un certain nombre de pays, dont la Géorgie, où une commission d'État sur les questions de migration a été spécialement créée pour élaborer un cadre national intégré de données sur les migrations faisant appel, dans un premier temps, à des recensements, des enquêtes et des données administratives. Ce travail permet de mesurer tous les indicateurs des objectifs de développement durable en ce qui concerne la ventilation des résultats par statut migratoire.

16. En plus de ces données traditionnelles, le projet exécuté en Géorgie a commencé à s'intéresser à de nouvelles sources et technologies pourvoyeuses de données et s'est associé avec l'autorité nationale de régulation de la téléphonie mobile qui a accès aux données issues de ces réseaux et en effectue le traitement. En collaboration avec des partenaires internationaux tels qu'Eurostat, l'Union internationale des télécommunications, l'Organisation internationale pour les migrations et Positium, la Division tente d'améliorer la mesure des déplacements de population à l'aide de données de positionnement mobile et de données issues des médias sociaux, pour établir des statistiques distinguant celles des mouvements migratoires, dont les migrations alternantes, de celles du tourisme.

17. À ce jour, la modernisation du calcul des indices de prix a mis l'accent sur les données saisies par lecture optique provenant des détaillants, l'objectif étant d'accroître l'utilisation effective de ces données en statistique officielle. L'équipe spéciale du Groupe de travail mondial chargée de ce domaine a l'intention de fournir un outil hébergé sur la plateforme mondiale, destiné à effectuer des travaux d'analyse, de suivi et de calcul d'indices, à partir des données de lecture optique historiques fournies par la société d'information, de données et de mesure Nielsen, qui s'accompagnera de supports éducatifs et de modules de formation sur l'utilisation de l'outil, de documents d'orientation méthodologiques et d'un catalogue de bonnes pratiques. Les données saisies par lecture optique offrent notamment la possibilité de modifier la pondération des dépenses au fil du temps. Sur la plateforme mondiale, les organismes nationaux de statistique ont désormais accès au code de méthode d'indices testés et peuvent s'exercer à calculer les indices à l'aide de différentes méthodologies applicables à des données d'entraînement.

18. Nielsen collabore avec le Groupe de travail mondial. Les données sont la spécialité de cette société qui recueille, enrichit et fournit des informations sur ce que les gens regardent, écoutent et achètent. Elle a également commencé à recueillir des données sur le commerce électronique à l'échelle mondiale, avec un taux de couverture de 76 % en 2018. Depuis un an, Nielsen et le Groupe de travail recherchent un terrain d'entente et sont en train de mettre au point des scénarios gagnant-gagnant dans leur coopération.

19. Début 2018, le Groupe de travail a constitué une équipe spéciale chargée des techniques de protection de la vie privée dirigée par l'Office for the National Statistics, qui devrait rédiger la section traitant de l'encodage du cadre de gestion des données applicable à la gouvernance des données et à la gestion de l'information. Cette équipe spéciale élaborera et proposera des principes, des politiques et des normes ouvertes pour l'encodage dans la plateforme mondiale. Ces règles établiront un usage déontologique des données qui répondra pleinement aux problèmes de sécurité et de confidentialité lorsqu'il s'agira d'élaborer des méthodes et des procédures de collecte, de traitement, de stockage et de présentation des données. Cela devrait notamment réduire les risques liés à la manipulation d'informations sensibles ou protégées.

20. Cybernetica, membre actif de l'équipe spéciale du Groupe de travail chargée des techniques de protection de la vie privée, a présenté ces techniques appliquées à l'utilisation des données saisies par lecture optique et issues des téléphones mobiles. On recourt à la cryptographie pour rendre impossible la re-identification des sujets des données et réduire le risque d'attaques internes sans nuire à l'exactitude des résultats. On utilise par exemple le calcul multipartite sécurisé et la cryptographie homomorphe, que Cybernetica a appliqués avec succès à des études de cas réalisées pour le Gouvernement estonien. L'anonymisation est une autre technique qui ajoute du bruit et rend la re-identification des sujets plus difficile, mais réduit aussi l'exactitude des résultats. La confidentialité différentielle et la k-anonymisation sont d'autres exemples de techniques.

III. Plateforme mondiale des Nations Unies et équipes spéciales du Groupe de travail mondial

21. Au cours de ses quatre années d'existence, le Groupe de travail mondial a réparti ses activités entre plusieurs équipes spéciales, un comité chargé de la plateforme

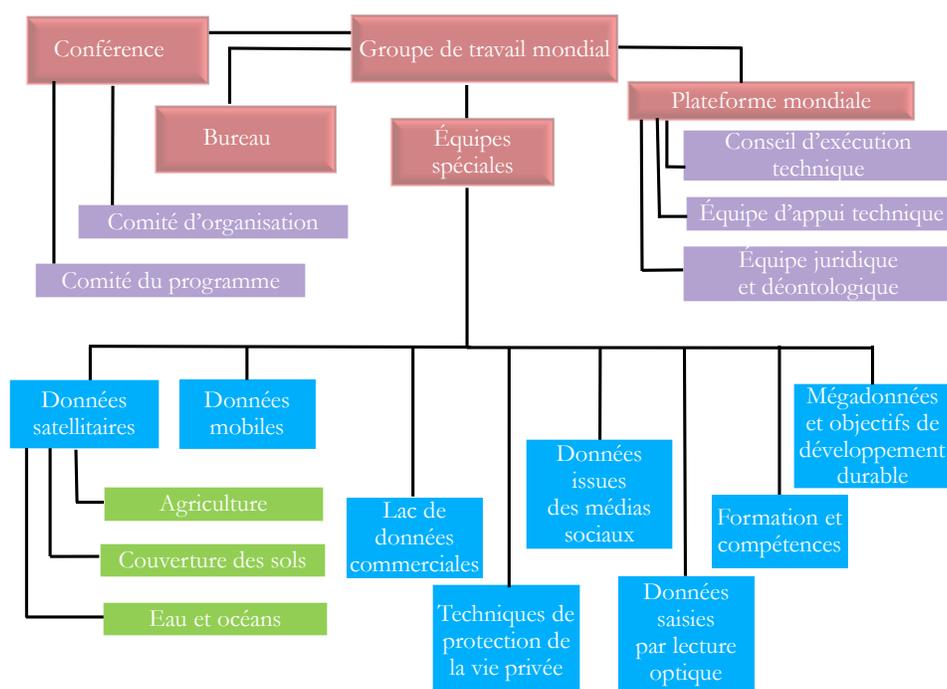
mondiale, des comités chargés de l'organisation des conférences internationales, ainsi que des domaines de travail et des projets connexes. Le Bureau du Groupe de travail mondial se réunit toutes les deux semaines afin de gérer et d'orienter les travaux des équipes spéciales et des comités. Chacune et chacun d'eux est dirigé par un institut qui fixe son propre calendrier de réunions et définit le programme de travail. On trouvera ci-dessous une vue d'ensemble de cet organigramme complexe.

A. Plateforme mondiale des Nations Unies

22. La Déclaration de Bogota (voir [E/CN.3/2018/8](#)) a précisé les éléments et buts de la plateforme mondiale, qui vise à faciliter le partage, l'échange et le développement de données, de métadonnées, de méthodes, de partenariats et d'outils d'apprentissage fiables. La plateforme fera partie d'un réseau interconnecté et fédéré de plateformes, reposera sur les pratiques exemplaires recensées dans le cadre des initiatives publiques et privées ayant trait aux mégadonnées, et offrira une infrastructure technologique favorisant l'innovation en matière de données au sein de la communauté des producteurs de statistiques officielles. La plateforme mondiale devrait appuyer le renforcement des capacités grâce à la constitution d'une bibliothèque de supports de formation, de méthodes et de logiciels fiables ainsi qu'à l'organisation d'ateliers sur la modernisation des statistiques officielles, le recours à de nouvelles sources de données (telles que les mégadonnées) et l'utilisation de nouveaux outils, services et techniques d'analyse. La plateforme, qui sera développée et mise à jour sous les auspices et la direction de la Commission de statistique, viendra appuyer les systèmes statistiques nationaux des pays développés et en développement.

Figure

Organigramme du Groupe de travail mondial



23. La plateforme mondiale comprend actuellement plusieurs services à l'état de prototypes tels que l'accès à Alibaba Cloud, à Amazon Web Services, à Google Cloud Platform et au nuage Azure de Microsoft, ainsi qu'à d'autres services permettant de collaborer à l'édition de codes, de publier des méthodes et d'analyser des données de localisation et d'observation de la Terre. Les utilisateurs de la plateforme mondiale peuvent rechercher, construire, déployer et consommer des algorithmes et développer davantage des méthodes statistiques à l'aide des principaux langages de programmation utilisés par les statisticiens (R, Python, Java et Scala). La plateforme peut également accueillir des modèles d'apprentissage automatique et publier les interfaces de programmation applicative (API) qui leur correspondent. Les partenaires du monde entier qui collaborent sur la plateforme peuvent utiliser les algorithmes depuis leur propre environnement en appelant les API. Ils auront également accès à plusieurs ensembles de données mondiales, tels que les données ADS-B sur le trafic aérien depuis juillet 2016, les données AIS sur le trafic maritime et les données de l'imagerie satellitaire commerciale à haute résolution.

B. Équipes spéciales du Groupe de travail mondial

24. Après la publication de son manuel, il y a un an, l'équipe spéciale sur l'imagerie satellitaire et les données géospatiales⁴ a mené avec succès un atelier de formation de cinq jours sur l'utilisation des données d'observation de la Terre à des fins statistiques ainsi que sur les méthodes d'estimation du rendement des cultures et les statistiques connexes à partir des données de l'imagerie satellitaire. Du fait que d'autres sujets sont venus s'ajouter aux statistiques agricoles, elle a créé trois domaines de travail, à savoir l'estimation de la production agricole, les statistiques de couverture des sols et d'occupation des sols, et les statistiques des écosystèmes liés à l'eau.

25. S'agissant du volet agricole, Statistique Canada a téléchargé sur la plateforme, aux fins de mise en commun et à l'essai, ses données satellitaires, ses données d'enquêtes sur les récoltes, ses données agroclimatiques, une partie du code source des rendements des cultures du Canada et la documentation d'appui. Cette approche collaborative de partage et de mise à l'essai aboutit à l'établissement de données, de méthodes et d'outils d'apprentissage fiables. Ce domaine de travail facilitera la mesure de l'indicateur 2.4.1 des objectifs de développement durable concernant la proportion des zones agricoles exploitées de manière productive et durable. Un petit nombre de projets ciblés recourant aux données satellitaires pour repérer des cultures spécifiques et déterminer leur rendement, au Canada, au Kenya et au Rwanda, sont également en préparation.

26. Les travaux effectués dans le domaine des statistiques de couverture des sols et d'occupation des sols ont particulièrement été axés sur le calcul des indicateurs des objectifs de développement durable 11.3.1 (ratio entre le taux d'utilisation des terres et le taux de croissance démographique) et 15.1.1 (proportion de la surface émergée totale couverte par des zones forestières). Ces indicateurs incluent la mesure de l'état et des modifications de l'occupation et de la couverture des sols au fil du temps. Les méthodes utilisées pour ces mesures sont validées par le Comité d'experts de la comptabilité environnementale et économique. Dans le cadre de ces travaux, l'exécution d'un projet portant sur l'évolution des tourbières au fil du temps est à l'étude.

⁴ Voir <https://unstats.un.org/bigdata/taskteams/satellite/>.

27. Un troisième domaine d'activité est celui de l'équipe spéciale sur l'imagerie satellitaire et les données géospatiales, qui s'intéresse à la variation de l'étendue des écosystèmes tributaires de l'eau (indicateur 6.6.1 des objectifs de développement durable). Comme il a été indiqué, on dispose d'une application utilisant les données satellitaires pour calculer l'étendue des eaux douces à l'échelle mondiale. Au niveau national, l'application peut être légèrement adaptée aux conditions locales, comme cela a été fait au Canada où, dans une large partie du pays, les nappes d'eau superficielles sont gelées durant l'hiver. Dans le cadre de ces travaux, un projet est à l'étude portant sur des bassins versants particuliers situés dans un ou plusieurs pays en développement.

28. En ce qui concerne les autres équipes spéciales du Groupe de travail, celle qui est chargée de l'utilisation des données issues des téléphones mobiles a terminé la première version complète de son manuel, qui présente en détail les sources de données, les méthodes, les modèles de partenariat et les applications destinées principalement au calcul de statistiques de tourisme. L'équipe s'attaquera bientôt au deuxième volume du manuel, qui comprendra des applications supplémentaires destinées à mesurer la mobilité humaine et traitera de la protection de la vie privée dès la conception, à l'instar du cadre méthodologique d'Eurostat mentionné précédemment. L'équipe s'est fixée comme objectif prioritaire d'exécuter un projet portant sur la mesure des migrations et des déplacements touristiques, ainsi que le calcul de statistiques connexes, en Géorgie. Il est prévu de tester les méthodes et les algorithmes développés à l'occasion de ce projet en Colombie, en Indonésie et en Italie.

29. L'équipe spéciale chargée de l'utilisation des données saisies par lecture optique dans le calcul des indices des prix à la consommation est en train de tester des méthodes statistiques et un code logiciel, principalement développés à partir d'applications libres, et a rassemblé les résultats de ce travail dans un manuel. Cette publication permettra à d'autres bureaux de statistique de tester les données saisies par lecture optique et d'évaluer leur utilité potentielle dans la production de statistiques, en sus des données issues de la capture de pages Web et de celles tirées d'enquêtes. Une autre évolution prometteuse est la collaboration avec Nielsen qui a mis certaines de ses données à la disposition de la communauté statistique mondiale. Ces données constituent une source de données mondiales, normalisées et donc comparables, ce qui permettrait la mise en commun de méthodes fiables en vue du calcul des indices des prix à la consommation.

30. Enfin, telle que présentée par le Groupe de travail lors de la journée portes ouvertes, une nouvelle équipe spéciale chargée des techniques de protection de la vie privée élabore des méthodes et des procédures applicables au traitement et à l'échange sécurisés d'informations sensibles ou protégées, effectués sur la plateforme du Groupe de travail entre partenaires de confiance. L'équipe doit formuler et proposer des principes et des politiques d'encodage, élaborés à partir de normes ouvertes et d'algorithmes libres. Un manuel est en préparation et devrait être disponible dans le courant de 2019.

IV. Étapes à venir

31. Le Groupe de travail mondial doit poursuivre l'élaboration d'un modèle d'activité durable pour la plateforme, conçue comme un environnement de recherche-développement collaboratif au service des statisticiens du monde entier. Les projets et les outils d'apprentissage fiables sont des tests de validation de l'intérêt et de la

viabilité de la plateforme. En 2019, le Groupe de travail organisera et participera à une série de manifestations pour montrer les progrès accomplis dans la réalisation de la plateforme et son état de préparation.

A. Modèle d'activité pour la plateforme mondiale des Nations Unies

32. Les principaux aspects du modèle d'activité de la plateforme sont indiqués ci-après. De plus amples détails seront fournis dans un complément d'information au présent rapport.

Personne morale

33. Une entité juridique est nécessaire en tant que support du fonctionnement général de la plateforme mondiale. La structure et la propriété de cette entité restent à trouver. L'entité devrait pouvoir solliciter et gérer des fonds et être en mesure d'assumer les risques liés à l'exploitation de la plateforme. Un modèle comme celui du Centre mondial de surveillance pour la conservation du Programme des Nations Unies pour l'environnement pourrait être une bonne solution.

Entité opérationnelle

34. Il est envisagé que les partenaires de confiance de la plateforme seront à la fois des fournisseurs et des utilisateurs de ses produits et services. Différentes parties des mêmes instituts participants peuvent être aussi bien des fournisseurs que des utilisateurs, les organismes de statistique officielle et leurs partenaires ayant une position particulière. Les règles d'engagement auxquelles sont soumis les différents types de partenaires restent à établir. La plateforme mondiale donnera accès, dans un premier temps, à des sources de données ouvertes, autorisant des accès plus ou moins libres à des données plus sensibles au fil du temps. Elle devrait bénéficier d'un appui à l'échelle mondiale, 24 heures sur 24. Le modèle pourrait être étendu à des centres régionaux pour mener des activités de codéveloppement et de renforcement des capacités.

Fonds

35. Moyennant des ressources importantes, pouvant provenir de sources philanthropiques ou de sources de financement du développement, la plateforme pourrait croître très rapidement. Les investisseurs types pourraient être, outre ces sources de financement du développement, des fondations ou de grandes entreprises du secteur technologique dans le cadre de leurs activités de mécénat.

B. Test de validation : les projets

36. Presque toutes les équipes spéciales ont commencé à travailler sur un ou plusieurs projets dont l'exécution sur la plateforme mondiale sera un test de validation. Les projets ci-après, qui en sont à divers stades de développement, sont à signaler :

a) Estimation du rendement des cultures à l'aide de données satellitaires. Un bon exemple de ce type de projet est le travail susmentionné accompli par Statistique Canada, qui a utilisé avec succès des données satellitaires pour estimer le rendement de 15 cultures vivrières. Des projets similaires sont à l'étude pour des pays africains ;

b) Mesure de la variation de la couverture et de l'occupation des sols. Un projet est en préparation qui mesurera l'étendue des tourbières ;

c) Mesure de l'étendue des écosystèmes liés à l'eau. Un projet visant à estimer l'étendue des nappes d'eau douce du Canada a démarré et pourrait être appliqué à des zones de delta spécifiques (par exemple le delta du Mékong) ;

d) Mesure de la mobilité humaine. Un projet a débuté en Géorgie, dont le but est de mesurer les déplacements imputables aux migrants, aux touristes, aux travailleurs saisonniers et à la population (de jour et de nuit), à partir des données issues des téléphones mobiles ;

e) Estimation de l'indice des prix à la consommation, à partir de données fournies par Nielsen. Les premiers essais de calcul de l'indice des prix à partir des données de Nielsen au Canada ont été effectués. D'autres essais impliquant un éventail de produits plus large et d'autres pays sont prévus pour 2019 ;

f) Analyse exploratoire de données à partir de celles relatives au commerce et au transport. Un lac de données comprenant des statistiques commerciales, les données AIS de trafic maritime et les données ADS-B de trafic aérien a été installé sur la plateforme, et les essais commenceront en 2019.

C. Test de validation : les outils d'apprentissage fiables

37. Les quatre ateliers de formation de deux jours organisés à Bogota en novembre 2017 ont montré que le Groupe de travail mondial était en mesure d'offrir un programme de formation sur l'utilisation des nouvelles sources de données pour la production de statistiques officielles. Un atelier régional de formation de cinq jours sur l'utilisation d'images satellitaires aux fins de calcul de statistiques agricoles a été organisé à Bangkok en juin 2018 pour la région de l'Asie et du Pacifique. Un programme similaire d'ateliers sera proposé en 2019, avec des sessions de formation plus courtes juste avant la cinquième Conférence internationale sur les mégadonnées et un atelier régional de cinq jours sur l'utilisation des données issues des téléphones mobiles en statistique officielle, qui se tiendra à Jakarta dans la quinzaine de juin 2019.

38. L'équipe spéciale chargée de la formation, des compétences et du renforcement des capacités est entrée dans une nouvelle phase d'activité avec de nouvelles offres, mettant davantage l'accent sur le développement des compétences dans un environnement en pleine évolution en matière de données. Statistics Poland a pris l'initiative sur cette question. L'Europe s'est fixée comme objectif, à long terme, de pourvoir l'ensemble du système statistique européen d'un grand nombre de diplômés en statistique experts en science des données. En 2020, cette science devrait faire partie intégrante de l'enseignement de la statistique officielle. Le programme de formation professionnelle du Groupe de travail mondial sera étroitement lié aux programmes existants dans le monde, notamment le Programme européen de formation statistique, qui offre des cours sur l'utilisation des sources de mégadonnées, comme l'analyse de texte appliquée aux médias sociaux et aux recherches sur le Web ou l'utilisation des données issues des téléphones mobiles en statistique officielle.

D. Manifestations

39. En 2019, le Groupe de travail mondial aura plusieurs occasions de présenter l'état d'avancement de la plateforme mondiale. Plusieurs manifestations seront organisées :

a) Une manifestation parallèle lors de la cinquantième session de la Commission de statistique, qui se tiendra à New York en mars 2019. Comme indiqué précédemment, le Groupe de travail publiera, en complément du présent rapport, un document d'information présentant les différentes options pour le modèle d'activité durable de la plateforme mondiale. La manifestation parallèle sera l'occasion de les expliquer et d'en débattre avec l'ensemble des statisticiens ;

b) La cinquième Conférence internationale sur les mégadonnées. Après l'Asie et le Pacifique, le Moyen-Orient, l'Europe et l'Amérique du Sud, l'Afrique accueillera cette conférence mondiale, donnant aux pays de ce continent l'occasion de montrer comment ils utilisent les nouvelles sources de données et les nouvelles technologies dans le calcul de leurs statistiques officielles. Comme indiqué précédemment, divers ateliers de formation portant sur des sujets liés aux mégadonnées seront proposés en marge de cette conférence ;

c) Une manifestation parallèle portant sur les mégadonnées et les nouvelles technologies se tiendra à Kuala Lumpur du 15 au 17 août 2019, à l'occasion du soixante-deuxième Congrès mondial de la statistique de l'Institut international de statistique. En coopération avec le comité d'organisation du Congrès, le Groupe de travail mondial a l'intention de proposer des exercices pratiques sur l'utilisation des données satellitaires, des données (d'entraînement) de positionnement mobile, des données (d'entraînement) issues des médias sociaux et des données (d'entraînement) saisies par lecture optique accessibles sur la plateforme mondiale. Le public ciblé est la communauté des statisticiens et des spécialistes des données des bureaux de statistique.

V. Suite à donner par la Commission de statistique

40. La Commission est invitée à prendre acte du présent rapport.

Annexe

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

Pays	Organisations
Allemagne	Banque africaine de développement
Arabie saoudite	Banque mondiale
Australie	Centre de statistique du Conseil de coopération des États arabes du Golfe
Bangladesh	Commission économique et sociale pour l'Asie et le Pacifique
Brésil	Commission économique pour l'Afrique
Cameroun	Commission économique pour l'Europe
Canada	Communauté des Caraïbes
Chine	Division de statistique
Colombie	Eurostat
Danemark	Fonds monétaire international
Égypte	Initiative Global Pulse de l'ONU
Émirats arabes unis	Institut de statistique pour l'Asie et le Pacifique
États-Unis d'Amérique	Organisation de coopération et de développement économiques
Géorgie	Organisation des Nations Unies pour l'alimentation et l'agriculture
Indonésie	Union internationale des télécommunications
Irlande	Union postale universelle
Italie	
Maroc	
Mexique	
Oman	
Pakistan	
Pays-Bas	
Philippines	
Pologne	
République de Corée	
République-Unie de Tanzanie	
Royaume-Uni de Grande-Bretagne et d'Irlande du Nord	
Suisse	