



---

**Commission économique pour l'Europe****Conférence des statisticiens européens****Soixante-dixième réunion plénière**

Genève, 20-22 juin 2022

Point 4 a) de l'ordre du jour provisoire

**Coordination des activités statistiques internationales  
dans la région de la Commission économique pour l'Europe :  
résultats des examens approfondis récemment réalisés  
par le Bureau de la Conférence des statisticiens européens****Examen approfondi de la collaboration avec les fournisseurs  
de données privés****Document établi par la Pologne (chef de file), le Canada,  
le Royaume-Uni, Eurostat et le Fonds monétaire international***Résumé*

On trouvera dans le présent document les résultats d'un examen approfondi des facteurs favorisant la coopération avec les fournisseurs de données privés et de ceux qui y font obstacle. Il comprend une analyse d'exemples de réussite vécus par des organismes nationaux de statistique et des organisations internationales ainsi que les enseignements tirés de leurs tentatives de collaboration restées infructueuses. Il permet aussi de mieux comprendre l'impact de la crise liée à la COVID-19 sur l'utilisation des données détenues par le secteur privé dans la statistique officielle. La dernière section résume les débats que le Bureau de la Conférence des statisticiens européens a tenus à sa réunion de février 2022.



## I. Résumé analytique

1. Il est avéré que les nouvelles sources de données, notamment celles détenues par le secteur privé, offrent un potentiel considérable pour ce qui est de renforcer et de compléter les statistiques officielles. Elles permettent souvent de produire dans de meilleurs délais des statistiques plus pertinentes, correspondant à des phénomènes socioéconomiques en évolution rapide et à des changements dynamiques des besoins des utilisateurs de données. Les autorités statistiques aux niveaux national et international se sont efforcées de faciliter l'accès aux données détenues par le secteur privé et leur utilisation dans la production de statistiques officielles. Compte tenu de l'importance de la collaboration avec les fournisseurs de données du secteur privé<sup>1</sup>, le Bureau de la Conférence des statisticiens européens (CSE) a décidé d'en faire le thème d'un examen approfondi.

2. Cet examen approfondi s'est appuyé sur une recherche documentaire réalisée à partir de notes d'orientation pertinentes, sur les résultats de recherches menées récemment au niveau international ainsi que sur des études de cas communiquées par 34 organismes nationaux de statistique et organisations internationales. La recherche documentaire a permis d'esquisser le contexte de l'examen et de dresser un panorama des activités des organisations internationales dans le domaine de l'élaboration de principes régissant l'accès aux données détenues par le secteur privé, la communication et la conclusion d'accords avec les diverses parties prenantes. Elle a également abordé la législation, le développement de lignes directrices et de manuels pour l'utilisation de ces données privées, les infrastructures informatiques communes et les initiatives de formation et de renforcement des compétences. Pour leur part, les études de cas ont permis d'approfondir des aspects qui pour l'heure n'ont fait l'objet d'aucune étude exhaustive, dont les facteurs favorisant la collaboration avec les détenteurs de données privés et notamment ceux qui ne sont pas d'ordre juridique, les principales causes des échecs de tentatives de coopération, ainsi que les bouleversements potentiels provoqués par la pandémie de COVID-19 en ce qui concerne l'accès à des données détenues par des entités privées et leur utilisation dans la statistique officielle.

3. En dépit de la diversité des expériences des organismes nationaux de statistique, les études de cas ont permis de recenser certains facteurs récurrents propices à la collaboration et à l'utilisation à des fins statistiques des données détenues par le secteur privé, dont : la confiance (perception du public, respect de la vie privée et confidentialité), la responsabilité sociale des entreprises, la communication, la compréhension et l'acceptation des intérêts de chaque partie et les efforts déployés pour parvenir à une situation gagnant-gagnant, la flexibilité de l'approche adoptée par les organismes nationaux de statistique quant au format des données ou à la technologie employée pour leur partage, l'instauration d'un dialogue permettant de comprendre les « boîtes noires » (qualité des données et méthodologie), le renforcement des capacités organisationnelles des organismes nationaux de statistique favorisant l'utilisation de nouvelles données (par exemple, l'infrastructure informatique, les compétences du personnel, etc.), la solidité du cadre juridique et l'ouverture à un éventuel dédommagement pour les frais engagés. Le rapport a cependant aussi démontré qu'il n'existait pas de solution universelle transférable d'un pays ou d'une organisation à l'autre.

4. L'examen a montré que la crise liée à la COVID-19 n'avait que moyennement renforcé la propension des entreprises privées à coopérer avec le secteur public et leur volonté de partager leurs données dans l'intérêt général. Certains organismes nationaux de statistique ont eu l'occasion de nouer des liens de coopération avec des détenteurs de données privés ou d'intensifier des relations déjà existantes, mais l'absence d'incitations juridiques ou financières fortes rend incertain l'accès à de nouvelles sources de données.

5. Les principaux résultats de la recherche documentaire et de l'analyse des études de cas ont permis de proposer deux séries de recommandations. Au niveau international, il est important de renforcer la coopération entre les équipes spéciales et les groupes d'experts qui examinent, sous l'égide des organisations internationales, la question de l'accès aux données détenues par le secteur privé. L'objectif doit être de parvenir à une spécialisation accrue, à une répartition claire des tâches et à une utilisation efficace des ressources humaines et

---

<sup>1</sup> Dans le rapport, les expressions « fournisseur de données privé » et « détenteur de données privé » sont considérées comme synonymes.

financières. Au niveau national, il est proposé de tirer parti des possibilités qu'ont les organismes publics de nouer des liens avec les fournisseurs de données privés, et d'investir dans des partenariats permettant un accès durable aux données et une meilleure compréhension des spécificités de ces dernières.

## II. Introduction

6. Le Bureau de la Conférence des statisticiens européens (CSE) examine régulièrement et de manière approfondie certains domaines statistiques choisis. L'objectif de ces examens est d'améliorer la coordination des activités statistiques dans la région de la CEE, d'aborder les problèmes émergents, et de favoriser les échanges de bonnes pratiques et l'apprentissage mutuel.

7. Le Bureau de la CSE a décidé de consacrer un examen approfondi à la collaboration avec les fournisseurs de données privés à sa réunion de février 2022. Les conclusions de l'examen par le Bureau seront présentées à la session plénière de la CSE en juin 2022. La coordination de l'examen est assurée par la Pologne, en coopération avec le Canada, le Mexique, le Royaume-Uni, Eurostat et le FMI.

8. L'objet du présent examen était de recenser les facteurs favorisant la coopération avec les fournisseurs de données privés et ceux qui y font obstacle. Le présent document aborde donc des exemples de réussite vécus par des organismes nationaux de statistique et des organisations internationales ainsi que les enseignements tirés de leurs tentatives de collaboration restées infructueuses. Il permet aussi de mieux comprendre l'impact de la crise liée à la COVID-19 sur l'utilisation dans la statistique officielle des données détenues par des entités privées. La recherche documentaire et l'analyse des études de cas ont abouti à la formulation de plusieurs recommandations prospectives.

## III. Domaine statistique visé

9. La révolution numérique a fait naître de nouvelles possibilités de produire et de collecter des données à partir de nouvelles sources. La nécessité de disposer dans de meilleurs délais de statistiques plus innovantes, apportant un éclairage sur des phénomènes socioéconomiques en évolution rapide, a été accentuée par la pandémie de COVID-19 et d'autres situations d'urgence (dont les catastrophes naturelles vécues dans certaines régions) et rend particulièrement cruciale l'exploitation de sources de données non traditionnelles. Pour relever ce défi, les autorités statistiques du monde entier ont cherché de nouvelles sources de données et tenté d'accéder aux données détenues par le secteur privé.

10. Ces données privées sont souvent appelées « mégadonnées », mais il convient de noter que ces termes ne sont pas synonymes. Toutes les données collectées par des entreprises privées ne sont pas des mégadonnées, et à l'inverse toutes les sources de mégadonnées ne sont pas privées. Cependant, une bonne part des mégadonnées est détenue par des entreprises privées, ce qui explique que certains exemples évoqués ci-dessous y fassent référence.

11. Le présent examen approfondi porte sur des données qui, pour la plupart, ne sont pas publiques et dont l'accès doit être autorisé par leurs détenteurs. Il va sans dire qu'une définition aussi générale couvre des types de données très différents, présentant des caractéristiques variées et impliquant des approches différentes en termes juridiques, organisationnels et techniques.

12. L'une des premières tentatives de classification des nouvelles sources de données pour la statistique officielle a été entreprise en 2013 par l'Équipe spéciale des mégadonnées (créée sous l'égide de la CSE) de la Commission économique pour l'Europe. La taxonomie qu'elle a élaborée porte sur les sources de mégadonnées, mais la plupart d'entre elles sont détenues par des entreprises privées :

- a) **Des réseaux sociaux** (informations d'origine humaine) ;
- b) **Des entreprises traditionnelles** (données gérées par l'intermédiaire de processus ; cette catégorie englobe les données administratives, mais aussi celles relatives aux transactions commerciales, aux opérations bancaires et aux inventaires, au commerce électronique et aux cartes de crédit) ;

c) **L'Internet des objets** (données générées par machine, notamment par des capteurs fixes, des capteurs mobiles (données de géolocalisation des téléphones mobiles et données satellitaires), et données provenant de systèmes informatiques)<sup>2</sup>.

13. En 2020, le Groupe du Système statistique européen (SSE) sur l'utilisation de données privées pour la statistique officielle a proposé une classification des sources de mégadonnées détenues par des entreprises privées, dans le but de faciliter les activités juridiques, techniques, pratiques et financières faisant intervenir des groupes de sources, et non pas les sources de données individuelles (ESSC 2020). Cette classification couvre cinq dimensions essentielles de l'utilisation des données d'entreprises à des fins statistiques : i) l'associabilité des données (les données sont-elles directement ou indirectement associées à une personne/entité économique ou non) ; ii) la confidentialité commerciale (impact potentiel sur des intérêts commerciaux) ; iii) l'accès aux données (dans les locaux du fournisseur de données, via une interface de programmation d'application (interface API) ou autre accès direct, ou nécessité de collecter les données) ; iv) les détenteurs de données d'envergure mondiale (a trait au caractère national ou international de l'entreprise et à l'approche qui en découle pour accéder à la source) ; et v) l'écosystème des données (la matrice de quatre types de situation où interviennent un ou plusieurs fournisseurs de données et où le phénomène peut être couvert ou non par une source alternative de données autres que des mégadonnées). Ce schéma de classification a permis de définir cinq grandes catégories de données :

a) **Les données personnelles intelligentes détenues par des entités privées** (par exemple, données des réseaux de téléphonie mobile, données de transactions financières) ;

b) **Les données personnelles intelligentes détenues par des organismes publics** (par exemple, les données collectées à partir des appareils personnels intelligents, les données de santé publique) ;

c) **Les compteurs intelligents et les données machine** (par exemple, données issues des compteurs électriques, des compteurs de mesure du trafic routier, des systèmes d'agriculture intelligente, des lecteurs de codes-barres) ;

d) **Les interactions avec les médias Web** (par exemple, Twitter, les systèmes de réservation électroniques) ;

e) **Les renseignements affichés sur le Web** (par exemple, les sites Web des entreprises, les portails consacrés à l'emploi).

14. Les expériences des organismes nationaux de statistique en matière d'utilisation de données détenues par le secteur privé sont très diverses. En 2015, le Groupe de travail mondial des Nations Unies sur l'utilisation des mégadonnées en statistique officielle (devenu par la suite le Comité d'experts en mégadonnées et sciences des données en statistique officielle) a mené une enquête mondiale sur les mégadonnées auprès de l'ensemble des organismes de statistique<sup>3</sup>. Les questionnaires remplis et transmis par 93 pays ont montré que les mégadonnées les plus utilisées étaient les données issues des lecteurs de codes-barres, des images satellitaires et celles extraites du Web. Les données recueillies à partir des médias sociaux et des téléphones portables sont moins utilisées pour diverses raisons liées notamment au respect de la vie privée et à la confidentialité. En outre, l'enquête a permis de dresser un aperçu détaillé de 115 projets consacrés aux mégadonnées menés par les organismes nationaux de statistique et Eurostat. Elle a révélé des différences dans les niveaux d'aptitude des organismes nationaux de statistique dans ce domaine, mais aussi un besoin impérieux et urgent de disposer d'orientations sur l'accès aux sources de mégadonnées, d'élaborer des normes et des cadres de qualité communs, et de renforcer les capacités du personnel. Dans les années qui ont suivi, plusieurs organisations internationales se sont efforcées de combler ces lacunes.

<sup>2</sup> <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data> [consulté le 13 décembre 2021].

<sup>3</sup> <https://unstats.un.org/bigdata/documents/reports/GWG%20Background%20document%20-%202016%20-6-Report-of-the-2015-Big-Data-Survey-E.pdf> [consulté le 13 décembre 2021].

## IV. Aperçu des activités statistiques internationales en la matière

15. De nombreuses organisations consacrent des travaux de recherche et d'autres activités à la question de l'accès aux données détenues par le secteur privé et de leur utilisation à des fins statistiques. En raison des innombrables équipes et groupes spéciaux traitant de ces questions sous les auspices d'organisations internationales, on s'est efforcé, dans le présent rapport, de regrouper leurs activités par domaines thématiques. Le rapport recense les domaines bénéficiant de la participation active de nombreux acteurs, ceux dont les responsabilités sont clairement divisées, et éventuellement les sujets ou les lacunes qui n'ont pas encore été suffisamment abordés. Il convient cependant de noter que toutes les activités n'ont pas été mentionnées et décrites en détail en raison des contraintes de longueur du document.

### A. Principes d'accès aux données détenues par le secteur privé

16. Les travaux entrepris par les organisations internationales ont permis l'élaboration de recommandations pour l'acquisition de données détenues par le secteur privé. Ces règles ne sont pas juridiquement contraignantes, elles sont à considérer par les organismes nationaux de statistique comme des directives pratiques à prendre en considération avant de solliciter l'accès à des données. Outre leur rôle d'appui pour l'accès, ces documents ont également une importance en termes de communication, insistant sur les objectifs d'intérêt public, le caractère hautement sensible de l'utilisation des données d'entreprises et la nécessité d'une approche responsable.

17. En 2015, le Comité d'experts de l'ONU en mégadonnées et sciences des données en statistique officielle a rédigé un projet de Principes d'accès aux sources de mégadonnées pour la statistique officielle. Par la suite, le document a été étendu, remanié et renommé en « Recommendations for Access to proprietary data » (Recommandations pour l'accès aux données couvertes par un droit de propriété), qui renvoie à trois aspects de l'accès aux données des secteurs privé et public : les aspects qualitatifs de la gestion des données couvertes par un droit de propriété, les problèmes de confidentialité et de protection, et enfin les accords de partenariat<sup>4</sup>. Une liste de principes directeurs à appliquer lors de l'accès aux données détenues par le secteur privé au niveau du SSE a été élaborée en 2020 par le Groupe sur l'utilisation de données privées pour la statistique officielle (ESSC 2020). En 2021, le Conseil de l'OCDE a adopté la Recommandation sur l'amélioration de l'accès aux données et leur partage, premier ensemble de principes et orientations sur les moyens, pour les pouvoirs publics, d'optimiser les avantages du partage des données tout en protégeant les droits des individus et des organisations (OECD 2021).

### B. Communication et conclusion d'accords avec les parties prenantes

18. Les organisations internationales se sont efforcées d'engager le dialogue avec un large éventail de parties prenantes, notamment les principaux acteurs économiques, sur l'importance des nouvelles sources de données pour la production de statistiques dans des délais plus courts à l'appui de la prise de décisions. Au niveau mondial, le Comité d'experts en mégadonnées et sciences des données en statistique officielle a coopéré avec plusieurs entreprises privées par l'intermédiaire des équipes spéciales en charge de différents groupes de sources de données. En 2021, il a formé une nouvelle équipe spécialement chargée d'étudier l'acquisition de données du secteur privé à l'échelle mondiale, son rôle étant de prendre contact avec des entreprises d'envergure mondiale, de négocier l'accès à leurs sources de données pour l'élaboration de statistiques nationales, régionales et mondiales, et d'accélérer la réalisation du Programme de développement durable à l'horizon 2030.

<sup>4</sup> <https://unstats.un.org/unsd/statcom/48th-session/documents/BG-3d-recommendations-for-access-to-proprietary-data-E.pdf> [consulté le 13 décembre 2021].

19. La Commission européenne a participé activement à des activités similaires. En 2018, elle a formé un groupe de 23 experts indépendants chargé d'étudier le partage de données entre les entreprises et les pouvoirs publics (B2G). Ce groupe a formulé des recommandations destinées à surmonter les obstacles entravant la collaboration en matière de partage de données entre les secteurs public et privé<sup>5</sup>. En mars 2020, la Commission a réussi à conclure un accord de coopération dans ce domaine avec quatre plateformes d'économie collaborative : Airbnb, Booking, le groupe Expedia et TripAdvisor. L'accord entre les sociétés privées et Eurostat a été signé pour permettre l'accès à des données agrégées relatives aux services d'hébergement de courte durée, d'une importance toute particulière pour les statistiques du tourisme actualisées<sup>6</sup> :

20. Actuellement, les organisations internationales unissent leurs forces pour conclure des accords avec des fournisseurs de données privés et favoriser ainsi l'exploitation de sources de données non traditionnelles à l'appui de leur recherche en développement international. L'initiative « Development Data Partnership », menée par la Banque mondiale en collaboration avec le FMI, la Banque interaméricaine de développement, l'OCDE et le PNUD<sup>7</sup>, est un parfait exemple de ce type d'accord. Ce partenariat a pour objectifs de favoriser l'accès aux données du secteur privé au bénéfice du développement international, d'optimiser les ressources en données scientifiques et techniques au sein des organisations membres, et de définir et défendre des principes communs et des bonnes pratiques pour une collaboration responsable entre public et privé dans le domaine du développement international. D'autres organisations internationales sont en discussion pour rejoindre l'initiative.

## C. Législation

21. Pour les membres de l'Union européenne, la législation communautaire a un impact fort sur l'accès aux données détenues par le secteur privé. Le Règlement général sur la protection des données (RGPD) et la législation ePrivacy (sur la vie privée et les communications électroniques) transposés dans le droit national constituent un cadre juridique solide assurant la protection des données personnelles dans les pays de l'UE. Actuellement, d'autres travaux d'envergure sont menés pour permettre l'utilisation durable à des fins statistiques des données détenues par le secteur privé. En 2021, le Parlement européen a adopté la Stratégie européenne pour les données, dont l'objectif est de « permettre à l'UE de devenir l'économie habile à tirer parti des données la plus attrayante, la plus sûre et la plus dynamique du monde, en donnant à l'Europe les moyens d'agir à l'aide de données permettant d'améliorer ses décisions et de rehausser les conditions de vie de tous ses citoyens ». Par la suite, la Commission a adopté la Directive sur les données ouvertes et a élaboré le Règlement sur la gouvernance européenne des données (qui devrait être adopté prochainement), ainsi que la loi sur les données (en préparation, publication de la proposition prévue en 2022). Cette loi se veut une incitation législative majeure en Europe et comprendra un chapitre sur le partage des données entre entreprises et pouvoirs publics dans l'intérêt général. Elle sera ensuite transposée dans les législations nationales. En outre, une révision de la législation sectorielle, notamment du Règlement n° 223/2009 relatif aux statistiques européennes, est en cours.

22. Les aspects juridiques de l'accès aux données détenues par le secteur privé ont été examinés au cours des réunions de 2020 et 2021 des experts de la CEE sur la modernisation de la législation statistique. En 2020, ces experts ont conclu à la nécessité d'une législation moderne pour conforter les relations de confiance établies sur la base des avantages mutuels de la collaboration. Des règles nationales ou internationales normalisant et harmonisant les principes du partage des données privées sont indispensables pour donner aux entreprises

<sup>5</sup> [b2g\\_data\\_sharing\\_report\\_hleg\\_0891ACFC-BF33-AA88-24DA996D50D4D220\\_64954.pdf](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-stay_accommodation_offered_via_online_collaborative_economy_platforms&oldid=548927)

[consulté le 13 décembre 2021, à l'adresse <https://digital-strategy.ec.europa.eu/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more>].

<sup>6</sup> [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-stay\\_accommodation\\_offered\\_via\\_online\\_collaborative\\_economy\\_platforms&oldid=548927](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-stay_accommodation_offered_via_online_collaborative_economy_platforms&oldid=548927) [consulté le 13 décembre 2021].

<sup>7</sup> <https://datapartnership.org/> [consulté le 30 décembre 2021].

l'assurance qu'elles appliquent les bonnes pratiques d'éthique et de respect de la vie privée<sup>8</sup>. En 2021, la réunion a conclu que les cadres juridiques nationaux actuellement en vigueur devaient être développés davantage pour soutenir l'accès aux données détenues par le secteur privé, mais que se posaient également des problèmes liés à l'application pratique de la loi, à l'organisation des processus commerciaux concernés et à la communication avec le grand public. Les capacités opérationnelles globales des organismes nationaux de statistique sont à renforcer pour leur permettre de surmonter efficacement tous les obstacles posés à l'accès aux données détenues par le secteur privé, notamment trouver le juste équilibre entre les différents aspects juridiques<sup>9</sup>.

#### **D. Lignes directrices et manuels pour l'utilisation de données détenues par le secteur privé**

23. Si l'accès à de nouvelles sources de données est une question particulièrement importante, l'intégration des données détenues par le secteur privé dans la production statistique officielle n'est pas un processus simple et soulève de nombreux problèmes de qualité. C'est pourquoi les organisations internationales ont tenté d'élaborer des lignes directrices et des manuels pour aider les organismes nationaux de statistique à surmonter les difficultés les plus courantes liées à des sources de données particulières. En 2013, le Groupe de haut niveau de la CEE sur la modernisation de la production et des services statistiques a mis en place une équipe spéciale qui a élaboré un cadre de qualité pour les « données massives »<sup>10</sup> et publié en 2014 un examen approfondi des données massives<sup>11</sup>. Au sein du Comité d'experts en mégadonnées et sciences des données en statistique officielle, plusieurs équipes spéciales ont été formées pour travailler sur des sources de données précises, notamment les données de téléphonie mobile, les données issues des lecteurs de codes-barres, les données d'imagerie satellitaire, les données de systèmes d'identification automatique (suivi des navires), ou sur des questions transversales, telles que les techniques de préservation de la vie privée, la formation, les compétences et renforcement des capacités, l'exploitation des mégadonnées pour la réalisation des objectifs de développement durable (ODD) et l'évaluation de l'accès aux transports dans les zones rurales. Ces équipes élaborent des méthodes, des lignes directrices et des manuels pratiques, et proposent aux organismes nationaux de statistique des actions de renforcement des capacités<sup>12</sup>.

24. Début 2021, la CEE a publié le « *Guide to Sharing Economic Data in Official Statistics* » (Guide pour la mise en commun des données économiques dans les statistiques officielles)<sup>13</sup>. Si ce guide fournit principalement des outils destinés à faciliter et améliorer le partage des données entre les producteurs de statistiques officielles, ses conseils en matière de communication avec les fournisseurs de données (chap. 5) et de conclusion d'accords pour l'utilisation des données et la mise en place d'une infrastructure d'échange sécurisée (chap. 6) pourraient se révéler utiles pour la collaboration avec les fournisseurs de données du secteur privé. Cette pertinence se retrouve dans l'analyse des principaux aspects de la mise en commun des données et des obstacles et moyens de faciliter le partage des données pour surmonter lesdits obstacles (chap. 3).

25. Le problème de la qualité des données a été évoqué dans de nombreux projets d'organisations internationales visant à faciliter le partage des connaissances au sein de la communauté statistique et à proposer des lignes directrices pour des enquêtes générales ou plus ciblées. Il convient de mentionner, entre autres, les projets multipartenaires ESSnet Big Data menés au sein du système statistique européen au cours des années 2016-2018 et 2018-2020. Ils ont fourni des résultats précieux concernant différentes sources de données

<sup>8</sup> <https://unece.org/info/events/event/348362> [consulté le 30 décembre 2021].

<sup>9</sup> <https://unece.org/info/events/event/355411> [consulté le 30 décembre 2021].

<sup>10</sup> [https://ec.europa.eu/eurostat/cros/system/files/Task%20Team%20Big%20Data%20Quality%20Framework\\_937\\_unblinded\\_v1.pdf](https://ec.europa.eu/eurostat/cros/system/files/Task%20Team%20Big%20Data%20Quality%20Framework_937_unblinded_v1.pdf) [consulté le 30 décembre 2021].

<sup>11</sup> [https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/7-Examen\\_approfondi\\_des\\_donn%C3%A9es\\_massives\\_FR.pdf](https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/7-Examen_approfondi_des_donn%C3%A9es_massives_FR.pdf) [consulté le 30 décembre 2021].

<sup>12</sup> <https://unstats.un.org/bigdata/task-teams/index.cshtml> [consulté le 13 décembre 2021].

<sup>13</sup> <https://unece.org/statistics/publications/guide-sharing-economic-data-official-statistics> [consulté le 30 décembre 2021].

ainsi que des sujets transversaux : offres d'emploi en ligne, caractéristiques des entreprises, énergie intelligente, suivi des navires, données sur les transactions financières, données d'observation de la Terre, données sur les réseaux mobiles, statistiques innovantes sur le tourisme, processus et architecture, méthodologie et qualité, et préparation de statistiques intelligentes<sup>14</sup>. Certains thèmes continuent d'être développés dans le cadre de projets actuels, tels que le réseau ESSnet Trusted Smart Statistics – Web Intelligence Network<sup>15</sup> ou ESSnet Smart Surveys<sup>16</sup>. Autre initiative pertinente méritant d'être citée : le projet d'apprentissage automatique lancé en 2019 par le Groupe de haut niveau de la CEE sur la modernisation de la production et des services statistiques dans le but de recenser et de relever certains défis communs liés à l'utilisation de l'apprentissage automatique dans la production statistique, d'une importance particulière dans le cas de nouvelles sources de données<sup>17</sup>. Le projet est suivi par le Groupe 2021 sur l'apprentissage en ligne, qui le prolongera en 2022.

## E. Infrastructure informatique commune

26. Les organisations internationales sont aussi particulièrement actives dans la mise en place d'infrastructures informatiques communes pour faciliter l'accès aux sources de mégadonnées, ainsi que dans la mise au point d'outils de traitement et d'analyse de ces données. Le Comité d'experts en mégadonnées et sciences des données en statistique officielle a créé la Plateforme mondiale des Nations Unies<sup>18</sup> (un écosystème de services en nuage doté de quatre hubs régionaux) dont le but est de promouvoir la collaboration internationale pour l'utilisation de nouvelles sources de données dans la statistique officielle et de mesurer les progrès accomplis dans la réalisation des objectifs de développement durable. Les attentes concernant les données, les outils et les services offerts sont en constante augmentation à mesure que la plateforme prend de l'ampleur. En outre, l'OCDE a commencé à investir dans un environnement informatique appelé Smart Data platform, composé de quatre centres fonctionnels chargés de collecter, structurer et stocker toutes les données provenant de sources classiques ou alternatives<sup>19</sup>. La pierre angulaire du projet est la solution open source SDMX intitulée .Stat Suite, développée avec la Communauté de collaboration sur les systèmes d'information statistique (SIS-CC)<sup>20</sup>. Eurostat mène quelques projets informatiques pour développer les centres de collecte, de traitement et d'analyse des données provenant de divers types de sources (Trusted Smart Statistics Centre), notamment les données de centres de renseignements sur le Web, d'enquêtes intelligentes fiables, d'opérateurs de réseaux mobiles ou encore des données en rapport avec les transports, la logistique et les systèmes intelligents.

## F. Formations et renforcement des compétences

27. Pour mieux préparer les institutions à l'utilisation de nouvelles sources de données, les organisations internationales développent des activités de formation et de renforcement des compétences destinées aux organismes nationaux de statistique. Dans ce contexte, l'équipe spéciale du Comité d'experts en mégadonnées et sciences des données en statistique officielle chargée de la formation et du développement des compétences et des capacités leur fournit entre autres : le cadre de compétences répertoriant un ensemble complet de compétences utiles pour ceux qui travaillent à l'acquisition et au traitement des mégadonnées, ainsi qu'une matrice de maturité des mégadonnées, outil d'auto-évaluation permettant aux organismes nationaux de déterminer leur niveau de « maturité organisationnelle » s'agissant de l'utilisation des mégadonnées dans la production statistique. D'autres outils, dont un programme de formation aux mégadonnées, un catalogue de formation aux mégadonnées et

<sup>14</sup> [https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1_en) [consulté le 30 décembre 2021].

<sup>15</sup> [https://ec.europa.eu/eurostat/cros/content/project-overview\\_en](https://ec.europa.eu/eurostat/cros/content/project-overview_en) [consulté le 30 décembre 2021].

<sup>16</sup> [https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys\\_en](https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys_en) [consulté le 30 décembre 2021].

<sup>17</sup> <https://statswiki.unece.org/display/ML/HLG-MOS+Machine+Learning+Project> [consulté le 30 décembre 2021].

<sup>18</sup> <https://unstats.un.org/bigdata/un-global-platform.cshtml> [consulté le 30 décembre 2021].

<sup>19</sup> OECD Statistics and Data Governance – Terms of Reference [consulté le 13 décembre 2021].

<sup>20</sup> Stat Suite | Statistical Information System Collaboration Community (siscc.org) [consulté le 13 décembre 2021].



un nouveau système de gestion de l'apprentissage<sup>21</sup>, sont en cours de développement. D'autres équipes spéciales du Comité d'experts en mégadonnées et sciences des données en statistique officielle participent activement à ces développements, fournissant du matériel d'apprentissage et dispensant des formations sur les sujets traités.

## V. Pratiques nationales

28. Les informations présentées ci-dessous synthétisent les réponses apportées par les organismes nationaux de statistique à l'enquête sur la coopération avec les fournisseurs de données du secteur privé. Cette enquête a été réalisée au cours des mois de novembre et décembre 2021 dans le but de recueillir des informations en soutien de cet examen approfondi. Trente-quatre organisations de 32 pays<sup>22</sup> ont répondu au questionnaire, dont 28 pays membres de la CEE et 4 pays extérieurs à la région<sup>23</sup>.

29. Les études de cas ont été classées en deux catégories : les expériences fructueuses d'utilisation de données détenues par le secteur privé dans la production statistique officielle ou expérimentale, et les tentatives de collaboration avec des fournisseurs de données du secteur privé qui n'ont pas abouti au résultat escompté.

30. D'après de nombreuses enquêtes antérieures (par exemple, l'enquête menée en préparation de la discussion plénière du CSE de 2019 consacrée aux nouvelles sources de données ou le questionnaire de l'ESS sur l'accès légal aux données détenues par le secteur privé au niveau national réalisé en 2020), les données privées le plus fréquemment utilisées pour les statistiques de prix par les organismes nationaux de statistique sont celles issues des lecteurs de codes-barres et extraites du Web. Il s'agit d'un exemple particulièrement réussi d'intégration de données détenues par le secteur privé dans la production statistique officielle et il est de ce fait parfaitement documenté. Le présent examen approfondi ayant pour objectif d'enrichir les connaissances actuelles, il ne traite pas des études de cas relevant de ce domaine particulier.

31. Il convient de noter par ailleurs que la collaboration fructueuse n'est pas l'apanage d'un groupe particulier de pays. Habituellement, les organismes nationaux de statistique qui sont parvenus à accéder à des données détenues par le secteur privé pour produire des statistiques nouvelles ou améliorées ont aussi connu un échec à un moment ou à un autre.

### A. Les succès

32. Selon l'enquête, 28 des 34 participants ont évoqué au moins un exemple fructueux de collaboration avec des fournisseurs de données du secteur privé, l'organisation ayant réussi à exploiter de nouvelles sources de données obtenues grâce à cette collaboration pour la production statistique officielle ou expérimentale<sup>24</sup>. Parmi les sources de données acquises auprès d'entreprises privées, les plus fréquemment mentionnées sont : les données de localisation des téléphones mobiles, les données de transactions des institutions bancaires et financières, les données collectées sur les sites Web des entreprises, les plateformes nationales d'hébergement ou de billetterie, et les données tarifaires des entreprises privées de vente au détail (ce dernier type de source ne relève pas du champ d'étude principal défini pour l'examen). Le tableau 1 présente les exemples de sources de données auxquelles les organismes nationaux ayant participé à l'enquête ont accédé avec succès<sup>25</sup>.

<sup>21</sup> <https://unstats.un.org/bigdata/task-teams/training/index.cshtml> [consulté le 13 décembre 2021].

<sup>22</sup> Trois agences américaines ont répondu.

<sup>23</sup> Il s'agit de l'Australie, du Chili, de l'Équateur et du Mexique.

<sup>24</sup> Les exemples peuvent avoir fait référence à des productions ponctuelles ou régulières, ainsi qu'à des projets de recherche visant à préparer une future production statistique.

<sup>25</sup> Les exemples présentés dans le tableau correspondent aux informations fournies dans les questionnaires. Ils ne reflètent pas l'activité globale liée à l'utilisation de données privées par les organismes nationaux de statistique qui ont participé à l'enquête.

Tableau 1  
**Exemples de sources de données obtenues par les organismes nationaux de statistique auprès d'entreprises privées**

<i>Type de données</i>	<i>Domaines ou indicateurs statistiques</i>	<i>Pays</i>
Données de localisation de téléphones mobiles	Statistiques du tourisme, indicateurs de mobilité de la population	Allemagne, Autriche, Canada, Espagne, Irlande, Lettonie, Mexique, Portugal, Russie, Tchéquie
Données de transactions provenant d'institutions bancaires ou financières	Indicateurs économiques, par exemple estimations du chiffre d'affaires des entreprises, de la consommation des ménages, du produit intérieur brut (PIB), mesures de productivité industrielle	Australie, États-Unis, Mexique, Norvège, Portugal, Royaume-Uni
Données collectées auprès des sites Web des entreprises	Indicateurs économiques et d'emploi, par exemple les entreprises innovantes, l'importance de l'Internet pour l'économie, des indicateurs de cybersécurité, des offres d'emploi en ligne	Pays-Bas, Suède
Plateformes nationales d'hébergement ou de billetterie	Statistiques de tourisme, statistiques culturelles	Estonie, Suisse
Opérateurs de réseaux électrique ou de transport (routes, chemins de fer, compagnies aériennes)	Statistiques sur la consommation électrique, mesures de la productivité industrielle, statistiques sur le tourisme	Bélarus, Estonie, États-Unis
Données recueillies auprès des entreprises et de leurs associations dans des secteurs particuliers	Mesures de la productivité industrielle	États-Unis, Mexique
Données recueillies auprès de sociétés d'étude de marché	Enquête sur l'état d'esprit des consommateurs, l'accès aux produits alimentaires dans les supermarchés et les grandes épiceries, les indicateurs environnementaux, les habitudes alimentaires des consommateurs pour les repas consommés à la maison et hors de la maison	États-Unis, Ukraine

*Source* : Réponses au questionnaire de l'examen approfondi.

33. Si les tentatives présentées dans les études de cas ont été couronnées de succès, il a parfois fallu beaucoup de temps pour les mener à bien. Dans d'autres cas, l'organisme national de statistique a tenté d'approcher tous les principaux fournisseurs de données mais n'a réussi à engager une collaboration qu'avec l'un d'entre eux. Dans certaines circonstances, la coopération a été déclenchée par un événement ou une crise (voir sect. VI) ou a été initiée de manière plutôt accidentelle (par exemple, à l'occasion de conférences, de séminaires, etc.). Les sections suivantes dressent un aperçu des principaux aspects d'une collaboration réussie, en reprenant la structure du questionnaire.

34. **Comment les organismes nationaux de statistique ont-ils pris contact avec les fournisseurs privés de données ?** L'approche des organismes nationaux était à chaque fois différente et dépendait de l'entreprise concernée (de son secteur, de sa taille, etc.), ainsi que des relations éventuellement déjà nouées. Souvent, les contacts avec la haute direction d'une entreprise ont été initiés par le directeur général ou d'autres cadres supérieurs. Parfois, les négociations ont été engagées au niveau des experts, notamment lorsque le contact était déjà établi. L'aval de la haute direction de l'organisme national de statistique a été un élément

décisif dans le processus de négociation, mais la contribution d'experts en données et de spécialistes informatiques capables de fournir des informations plus détaillées sur les questions méthodologiques ou les systèmes de sécurité informatique a souvent été déterminante pour le succès des négociations. Le lobbying au sein des instances pertinentes a également été jugé important.

35. Les études de cas ont mis en évidence plusieurs autres facteurs importants dans les négociations avec les détenteurs de données privés : indiquer clairement l'objectif de l'utilisation des données, en insistant particulièrement sur le fait que celui-ci sert l'intérêt général, respecter le secret des affaires, s'employer à limiter au maximum la charge de travail pour l'entreprise et trouver des incitations convaincantes (pour parvenir à une situation gagnant-gagnant).

36. **En quoi la législation a-t-elle eu un impact sur l'accès aux données détenues par le secteur privé ?** Dans la plupart des pays, la législation nationale ne prévoit ni n'exclut l'accès aux données détenues par le secteur privé. Dans certains, la législation nationale a facilité l'accès à ces données de deux manières : en instaurant directement une obligation pour les entreprises privées de partager certaines données ou en fournissant une base solide de négociation avec les entreprises (cadres et exigences clairs comprenant, par exemple, des évaluations coûts-avantages). L'un des organismes nationaux de statistique a déclaré pouvoir exiger des données d'entreprises privées (autres que des petites entreprises ou microentreprises) en leur adressant une notification. Cependant, il ne recourait à cette solution qu'en dernier recours, après avoir épuisé toutes les autres formes de collaboration. En même temps, l'organisme a constaté que certains fournisseurs demandaient une notification officielle de manière à faire du partage des données une obligation légale. Indépendamment des autorisations légales, l'accès aux données détenues par le secteur privé était généralement obtenu sur une base volontaire et régi par des accords bilatéraux. Dans un cas, l'organisme national de statistique avait la capacité juridique de contraindre les entreprises privées à communiquer les données, mais en fait il leur achetait ces données.

37. En raison de la législation en matière de protection des données en vigueur aux niveaux européen (c'est-à-dire le RGPD) ou national, les entreprises privées ont de plus en plus conscience du risque lié au respect de la vie privée. Un participant à l'enquête a répondu qu'une société de téléphonie mobile avait demandé à ses clients la permission d'exploiter les données à des fins statistiques. Une bonne pratique partagée par un autre organisme national de statistique consistait à contacter les propriétaires des sites Web et à demander leur aval avant de mettre en œuvre des techniques d'extraction de données du Web (alors même que les conditions d'utilisation de robots d'extraction stipulées sur le site Web étaient satisfaites).

38. **Quels modèles de coopération ont été appliqués ?** Des modèles de coopération très divers ont été appliqués par les organismes nationaux de statistique, notamment en fonction des sources de données. Les organisations ont généralement acquis des données agrégées (dont des données de téléphones mobiles, des données de transactions provenant d'institutions financières, des données relatives aux hébergements de courte durée de plateformes en ligne), et parfois des données pseudonymisées (par exemple, les métadonnées téléphoniques du réseau mobile). Seuls deux organismes nationaux de statistique ont déclaré avoir accès aux données brutes (ou en avoir connaissance)<sup>26</sup>. S'agissant des données des téléphones portables ou des caisses enregistreuses, un tiers intervenait très souvent dans le processus d'acquisition et de traitement des données. Dans la plupart des cas, les données ont été obtenues gratuitement, mais certaines entreprises escomptaient une rémunération pour le partage ou la préparation et le traitement des données. Les exemples de modèles de collaboration sont brièvement décrits ci-dessous :

a) **Modèle de partenariat public-privé (PPP).** L'organisme national de statistique, sur la base d'arrangements institutionnels avec l'Association de l'industrie automobile, l'Association nationale des producteurs d'autobus, de camions et de tracteurs et chacune des entreprises impliquées, a lancé les registres administratifs de l'industrie

<sup>26</sup> Un organisme national de statistique a décidé de produire les registres administratifs de l'industrie automobile des véhicules légers et des poids lourds à partir des données brutes fournies par les entreprises, alors qu'un autre organisme a pu accéder aux bases de données de détenteurs de données privés.

automobile des véhicules légers et des poids lourds. Ce partenariat public-privé a pour objectif de collecter des informations brutes, puis de traiter et de diffuser des données plus fines sur l'industrie automobile<sup>27</sup> ;

b) **Modèle basé sur le détachement.** L'organisme national de statistique collabore avec l'entreprise privée en détachant en son sein l'un de ses fonctionnaires chargé d'aider à la mise en œuvre de la méthodologie et à l'analyse des données. La personne détachée peut partager régulièrement avec ses collègues de l'organisme les connaissances qu'elle a acquises mais sans pour autant partager les données elles-mêmes. Le détachement a facilité l'établissement de relations de partage de données entre l'organisme et l'entreprise ;

c) **Collecte de données externalisée.** L'organisme national de statistique coopère avec une entreprise spécialisée qui se charge d'extraire de manière régulière les données de sites Web, d'en dériver les caractéristiques standardisées et de lui communiquer ces données qui viennent utilement compléter celles obtenues par les enquêtes auprès des entreprises, les registres administratifs (par exemple, ceux des services fiscaux) et d'autres sources disponibles. Les données sont obtenues gratuitement. Un autre organisme national de statistique recueille des données collectées et traitées par une entreprise privée dans le cadre de sa propre enquête sur l'état d'esprit des consommateurs. Selon les termes du contrat conclu, les données appartiennent au fournisseur de données et leur propriété n'est pas transférée à l'organisme national de statistique. La méthodologie de recherche et l'ensemble des variables fournies sont fixés dans l'accord ;

d) **Spécification convenue de l'ensemble de données.** L'organisme national de statistique reçoit des ensembles de données strictement définis de la part du fournisseur de données (un tiers), qui traite les données du réseau mobile. La coopération est basée sur un accord régissant les aspects méthodologiques et techniques de l'échange, ainsi que la rémunération financière.

39. **Quelle infrastructure informatique a été utilisée pour obtenir et traiter les données ? À qui appartient-elle ?** Dans la majorité des cas, il n'a pas été nécessaire d'adapter l'infrastructure de l'organisme national de statistique pour collecter et traiter les données provenant des entreprises privées. Cependant, plusieurs organismes ont choisi d'investir dans un nouvel environnement de données sécurisé en vue de traiter des sources de mégadonnées (par exemple, Scaled Analytic Data Environment déployé via la technologie Cloud d'Amazon Web Services). Dans le cas des données d'un opérateur de réseau mobile, les solutions les plus fréquemment mises en œuvre reposaient sur un traitement initial, voire le traitement intégral réalisé par l'opérateur ou un tiers. Lorsque les données étaient transférées aux serveurs de l'organisme national de statistique, elles étaient soumises à des procédures de sécurité mises en œuvre par l'organisme (telles que les limitations d'accès, le cryptage, la pseudonymisation, etc.).

40. **Quels sont les problèmes de qualité des données qui ont été rencontrés ?** L'un des problèmes les plus fréquents auxquels ont été confrontés les organismes nationaux de statistique participant à l'examen a été celui de la couverture. Il résultait d'une part du défaut d'accès aux données de l'ensemble des principaux fournisseurs de données sur le marché, par exemple les opérateurs de réseau mobile ou les institutions financières, et d'autre part de la spécificité des données qui ne couvraient pas toujours l'ensemble de la population. Concernant les sources de mégadonnées, l'un des problèmes évoqués était la multiplicité des sources, avec pour corollaire le dilemme du choix entre une ou plusieurs sources. Par la suite, la nécessité d'élaborer une méthode d'uniformisation des données et de définir un cadre conceptuel commun de méthodes de collecte et de traitement des données a été soulignée.

41. Un autre ensemble de difficultés avait trait à la non-concordance entre les classifications employées par un détenteur privé de données et les classifications statistiques. Dans certains cas, cependant, il a été possible de coopérer avec les entreprises privées et d'adapter la classification à des fins statistiques.

<sup>27</sup> <https://en.www.inegi.org.mx/datosprimarios/iavl/> [consulté le 30 décembre 2021].

42. D'autres problèmes sont apparus, entre autres : les différences ou erreurs d'encodage des données, les modifications apportées au fil du temps, le manque ou l'obsolescence de certaines données, la couverture géographique limitée ou le niveau de détail insuffisant des données, le manque de cohérence entre les différents niveaux d'information, etc.

43. En raison de l'existence de « boîtes noires » et d'incertitudes quant à la qualité des données, l'intégration de nouvelles sources de données dans la production régulière s'avère délicate. L'une des approches proposées a été de maintenir le contact avec les entreprises privées et d'essayer de résoudre tous ces problèmes afin de trouver des solutions satisfaisantes. L'alternative consistait à accepter une qualité de données légèrement inférieure mais de pouvoir publier des données plus fréquemment ou concernant de phénomènes nouveaux.

44. Les études de cas ont permis de recenser plusieurs facteurs incitant les entreprises à entrer en collaboration avec les organismes nationaux de statistique et facilitant l'utilisation des données détenues par le secteur privé :

a) **La confiance** – La réputation de l'organisme national de statistique en tant qu'institution digne de confiance, fournissant des services publics précieux et appliquant des exigences légales strictes en matière de protection des données ;

b) **La responsabilité sociale** – Les programmes de responsabilité sociale des entreprises privées et leur contribution à la création d'un bien public ;

c) **La communication, la compréhension et la prise en compte des intérêts de chaque partie** – Une bonne préparation avant tout rapport avec l'entreprise, l'établissement et le maintien de contacts de haut niveau, des relations directes et personnelles, ainsi qu'une communication ouverte et une compréhension mutuelle ont été déterminants dans les négociations avec les détenteurs de données privés. Il s'agit d'expliquer l'objectif de la demande de données et de garantir leur utilisation à des fins exclusivement statistiques. La préservation de l'avantage concurrentiel étant fondamental pour les entreprises, il convient d'insister dans la communication bilatérale sur le fait que l'expérience de l'organisme national de statistique en matière de collecte et de traitement des données garantit la protection des données et ne remet nullement en cause le secret des affaires de l'entreprise ;

d) **Une situation gagnant-gagnant** – La stimulation doit être suffisante pour inciter l'entreprise privée à s'engager dans une coopération avec l'organisme national de statistique (par exemple, la garantie que la demande n'imposera pas une lourde charge de travail à l'entreprise, un dédommagement pour les frais engagés, la réduction de la charge administrative, un échange de données, etc.) ;

e) **Une approche flexible** – L'acceptation des données quel qu'en soit le format, le transfert des données à l'aide d'une technologie sécurisée que l'entreprise utilise et privilégie, ainsi que la limitation des demandes aux données les plus importantes facilitent considérablement les négociations ;

f) **La capacité organisationnelle de l'organisme national de statistique** – Il faut une « faim d'innovation », un personnel qualifié, une bonne coopération interne transversale, l'application de règles de respect de la vie privée, et des moyens appropriés s'agissant des solutions informatiques, des méthodes de travail et des capacités de gestion des achats ;

g) **Un cadre juridique** – Une base législative solide soutenant l'organisme national de statistique dans les négociations.

## B. Enseignements tirés des tentatives infructueuses de collaboration avec les détenteurs de données privés

45. Selon l'enquête, 20 des 34 participants ayant retourné le questionnaire ont évoqué au moins une tentative infructueuse de collaboration avec un fournisseur de données privé<sup>28</sup>. Plus de la moitié de ces échecs était en rapport avec des opérateurs de réseau mobile et l'utilisation des données de téléphonie mobile pour des statistiques officielles. Ce type de données semble particulièrement intéressant pour la production de statistiques fiables et actualisées, notamment dans les secteurs de la démographie, des migrations, du tourisme et de la mobilité. Cependant, l'expérience de nombreux pays montre qu'accéder aux données de téléphonie mobile et les utiliser régulièrement pour la production de statistiques n'est pas chose aisée.

46. Les études de cas ont mis en évidence un large éventail d'obstacles. Certains étaient liés à la perception sociale d'une telle coopération, d'autres à des conditions internes, d'autres encore à la qualité insatisfaisante des données. Les facteurs suivants ont été le plus fréquemment mentionnés comme ayant contribué à l'échec :

a) **Le manque de motivation des entreprises privées à coopérer** – Les entreprises ou les tiers d'envergure mondiale propriétaires de données telles que celles issues de la téléphonie mobile, des caisses enregistreuses ou des réseaux électriques ne sont pas enclins à coopérer avec l'organisme national de statistique de tel ou tel pays. Par ailleurs, les représentations locales de ces entreprises n'ont pas le pouvoir de décider de questions de nature stratégique, telles que le partage des données ;

b) **Des préoccupations concernant le respect de la vie privée et la perception du public** – Les entreprises sont très prudentes en matière de confidentialité des données. La simple perception d'un risque pour le respect de la vie privée qui serait susceptible de provoquer l'indignation de l'opinion publique peut les décourager de partager des données ;

c) **Des inquiétudes quant à une éventuelle divulgation de secrets commerciaux** – La protection des secrets commerciaux ou des informations confidentielles susceptibles d'avoir un impact sur la compétitivité de l'entreprise est un sujet particulièrement sensible. Les entreprises redoutent la divulgation de données commerciales stratégiques à leurs concurrents, d'où leur réticence à s'engager dans une collaboration de partage de données ;

d) **Le dédommagement financier** – Du fait des modèles commerciaux adoptés, certaines entreprises privées exigent une rémunération pour le partage de leurs données, une attitude qui pose problème à de nombreux organismes nationaux de statistique en raison des contraintes budgétaires auxquelles ils sont soumis ;

e) **L'absence de législation solide ouvrant la voie à une interprétation restrictive des règles existantes** – Les autorités chargées de la protection des données personnelles ou les services juridiques ont tendance à interpréter les règles juridiques relatives à l'acquisition de données privées de manière extrêmement restrictive, ce qui empêche toute collaboration avec les entreprises. Par ailleurs, les cadres juridiques ne sont pas assez solides pour créer des conditions favorables à cette collaboration ;

f) **Le caractère incomplet des données et leur peu d'utilité à des fins statistiques** – Sous-dénombrement ou surdénombrement des données, portée des variables obtenues auprès des entreprises privées (pas toujours utiles à des fins statistiques) ;

g) **Les « boîtes noires »** – Visibilité inadéquate sur les données brutes ou microagrégées, ou sur les métadonnées, et évaluations insuffisantes de la qualité des données, faisant obstacle à l'emploi de nouvelles sources de données dans la production statistique officielle ;

<sup>28</sup> Il peut s'agir d'une tentative ponctuelle ou régulière, ou encore de projets de recherche visant à préparer le terrain pour une production statistique future.

h) **Autres facteurs** – Le fournisseur de données s’est retiré du marché ; l’entreprise privée a modifié sa structure organisationnelle, ce qui empêche la poursuite de la coopération ; l’organisme national de statistique ne disposait pas des ressources nécessaires (ressources humaines, compétences ou infrastructure informatique adéquate) pour traiter les données.

## VI. Impact de la crise liée à la COVID-19

47. Près de la moitié des organismes nationaux de statistique interrogés ont déclaré que la pandémie de COVID-19 leur avait ouvert de nouvelles voies d’accès aux données détenues par le secteur privé. Selon eux, la crise sanitaire a amené les entreprises à s’ouvrir davantage au partage des données, ou a du moins facilité la communication en la matière. Cela étant, si les entreprises se sont montrées plus disposées à coopérer avec le secteur public, elles préféreraient nouer une collaboration non pas avec l’organisme national de statistique, mais plutôt directement avec les autorités nationales, car cela était plus valorisant pour leur image. Pendant la pandémie, quelques organismes nationaux de statistique ont commencé à coopérer avec les opérateurs de réseau mobile pour réaliser des analyses ponctuelles, à petite échelle sur une zone limitée, du comportement de la population en termes de mobilité. Certains ont obtenu les données d’institutions financières et ont ainsi pu estimer le chiffre d’affaires des entreprises, la consommation des ménages et le produit intérieur brut (PIB), et contribuer au suivi des tendances économiques. Dans quelques cas, la crise liée à la COVID-19 a engendré un renforcement de la coopération avec les chaînes de supermarchés, les enquêtes en face à face assistées par ordinateur (CAPI) étant jugées trop risquées.

48. La situation sanitaire a permis à des organismes nationaux de statistique d’initier une coopération avec des entreprises du secteur privé ou d’intensifier des relations existantes. À titre d’exemple, certains ont obtenu d’entreprises l’autorisation (généralement temporaire) de transmettre des données avec une fréquence accrue, d’affiner le degré de détail ou d’employer un catalogue de variables plus étendu, afin de soutenir la production rapide de statistiques ponctuelles liées à la crise.

49. Quelques organismes nationaux de statistique ont souligné que la propension des entreprises à œuvrer pour le bien public constituait une motivation au partage des données. Pour les entreprises privées, les avantages directs tirés de statistiques officielles de qualité et à jour étaient un autre facteur incitatif. Des entreprises se sont abstenues de réclamer un remboursement des frais engagés ou le paiement d’honoraires, mais cette approche était loin d’être universelle et aucune évolution majeure vers une plus grande motivation des entreprises à coopérer ou l’abandon des compensations financières n’a été signalée en marge des changements positifs.

50. Les études de cas n’ont pas mis en lumière d’incidence de la crise liée à la COVID-19 sur les règles juridiques relatives à l’accès aux données détenues par le secteur privé. Tous les cas de collaboration rapportés ont été régis par des accords bilatéraux. Parallèlement, certains organismes nationaux de statistique ont fait remarquer que l’urgence de l’accès à ces données s’était estompée à mesure que la pandémie « s’installait » et que les entreprises commençaient à se désengager. Ils espéraient l’adoption de mesures législatives pour assurer l’accès aux données des entreprises.

51. En réponse aux nouveaux défis liés à l’utilisation ou à la tentative d’utilisation de nouvelles sources de données, certains organismes nationaux de statistique ont procédé à des aménagements d’ordre organisationnel, en créant par exemple une équipe spécialisée, un groupe de travail interne ou en réaffectant des ressources humaines aux tâches en rapport avec la COVID-19. Quelques organismes ont modifié leur infrastructure informatique, développant des interfaces API pour le téléchargement des données quotidiennes des réseaux mobiles. Certains se sont tournés vers la technologie basée sur le Cloud pour acquérir, stocker et traiter des données à grande échelle, à haute fréquence et à diverses fins statistiques, ou ont mis en place une structure de gestion plus agile et personnalisée. D’autres encore ont instauré des changements plus normalisés, notamment en redéfinissant les rôles et les responsabilités en matière d’acquisition et de gouvernance des données, de gestion des relations avec les fournisseurs, ou en adoptant des politiques de financement claires pour l’acquisition de données stratégiques et les solutions informatiques connexes.

52. Tous les organismes nationaux de statistique ont affirmé leur volonté de préserver la coopération instaurée avec les entreprises privées. Ils s'efforceront par ailleurs de l'élargir et de nouer des contacts avec de nouvelles entreprises. Cependant, en l'absence d'une obligation formelle (à quelques exceptions près<sup>29</sup>), la situation reste aléatoire et la collaboration, qui prend parfois fin pour des raisons parfois peu compréhensibles, ne pourra être pérennisée qu'avec la bonne volonté de leurs interlocuteurs commerciaux. En outre, pour les échanges sujets à une rémunération financière, les ressources budgétaires de l'organisme national de statistique seront déterminantes pour la poursuite de l'acquisition de données.

## VII. Problèmes et enjeux

### A. Doubles emplois et absence de répartition claire des tâches au niveau international

53. L'examen des activités internationales dans le domaine de la collaboration avec les fournisseurs de données du secteur privé montre que les organisations internationales ont déployé des efforts substantiels dans ce domaine. Mais en l'absence d'une répartition claire des tâches et de synergie, certains efforts semblent faire double emploi (Biancotti et al., 2021). Cette situation peut se traduire par une allocation inefficace des ressources, tant humaines que financières. Elle peut aussi accroître la charge des entreprises privées, si celles-ci sont invitées à s'engager dans de multiples groupes de travail ou collaborations. En outre, les mandats des innombrables forums intervenant en la matière risquent de ne pas être véritablement clairs pour les entreprises et les organismes nationaux de statistique. Les résultats des travaux de toutes les parties concernées seraient plus reconnaissables et compréhensibles pour les utilisateurs potentiels si ces actions étaient davantage harmonisées. Il convient par ailleurs de souligner que les premières tentatives de mutualisation des forces ont déjà été menées, dont un parfait exemple est l'initiative Development Data Partnership, qui regroupe diverses organisations internationales.

### B. Outils législatifs limités à la disposition des organismes nationaux de statistique

54. La législation nationale de la plupart des pays ayant répondu à l'enquête ne favorise ni n'entrave l'accès aux données détenues par le secteur privé. Cependant, l'absence de base juridique solide est perçue par de nombreux organismes nationaux de statistique comme un obstacle majeur au dialogue avec les fournisseurs de données privés et à l'établissement d'un accès pérenne à leurs données à des fins statistiques. L'accès aux données privées est généralement obtenu sur une base volontaire et régi par des accords de coopération bilatéraux.

55. Le principal défi à relever est le peu de mesures incitatives à caractère permanent destinées aux détenteurs de données privés. Sans base législative solide, les organismes nationaux de statistique recourent principalement à des incitations douces, notamment l'évocation de l'intérêt général, les possibilités de produire plus rapidement des indicateurs économiques qui sont utiles pour l'entreprise, ou encore la non-augmentation, voire la réduction de la charge de travail des entreprises. Une autre approche consiste à proposer une rémunération financière, mais elle suppose souvent des règles de financement plus détaillées pour les acquisitions de données stratégiques.

<sup>29</sup> Une telle obligation existe par exemple en Norvège (la transmission des données liées aux transactions par carte de débit est sécurisée grâce à une nouvelle loi sur les statistiques introduite en 2019 et officialisée par une décision sur l'obligation de fournir des informations en juin 2021), au Royaume-Uni (l'organisme national de statistique peut exiger des données de la part d'entreprises privées, autres que des petites entreprises ou microentreprises, en leur adressant une notification), au Canada (Statistique Canada a la capacité législative d'imposer aux entreprises privées l'obligation de fournir leurs données) et en Serbie (un article prévoyant l'utilisation des données des entreprises privées à des fins statistiques a été ajouté au nouveau projet de loi relative aux statistiques).



56. Il n'existe pas de solution unique et transposable d'un pays à un autre. Des tentatives infructueuses ont été enregistrées même dans les pays où une coopération a été établie entre les organismes nationaux de statistique et les entreprises privées, et où des données privées ont été employées à des fins statistiques. En outre, les partenariats volontaires présentent un risque pour la pérennité de l'accès aux données et rendent incertaine la fourniture de statistiques fondées sur des données privées.

### C. Évaluations insuffisantes de la qualité des données

57. Le défaut d'accès aux données brutes ou micro-agrégées et la connaissance insuffisante de la qualité des données font du passage de la production expérimentale à la production officielle de statistiques un processus long et complexe. Seule une coopération étroite et franche entre l'organisme national de statistique et les entreprises privées peut résoudre bon nombre des problèmes de qualité associés à ce type de données. Le défi à relever est celui du renforcement de la coopération entre les experts de l'organisme national et les entreprises privées en matière de traitement des données, de méthodologie, de développement d'algorithmes et d'évaluation de la qualité des données. Le détachement temporaire de fonctionnaires de l'organisme national au sein de l'entreprise privée, dans le but de mieux comprendre la spécificité des données, peut faciliter la résolution des problèmes qualitatifs et de ce fait accélérer la production de statistiques officielles basées sur (ou enrichies par) des données détenues par le secteur privé.

### D. Perpétuer la coopération initiée à l'occasion de la crise de la COVID-19

58. La pandémie de COVID-19 a fait que les entreprises privées ont été dans une certaine mesure plus disposées à coopérer avec le secteur public et à lui fournir des données récentes et plus détaillées pour surveiller l'impact de la crise. En outre, certains organismes nationaux de statistique ont procédé à des aménagements en termes d'infrastructure informatique ou de gestion agile des ressources humaines et autres. Bien que l'examen ne se soit pas axé sur les compétences et le renforcement des capacités en tant que tels, il montre clairement que le partage des connaissances et des expériences (fructueuses ou non) entre les organismes nationaux de statistique peut jouer un rôle majeur en favorisant le recours à de nouvelles sources de données dans les statistiques officielles. Il en va ainsi des données issues des lecteurs de codes-barres et des données extraites du Web pour établir des statistiques de prix, dont l'accès a été sensiblement amélioré par la pandémie de COVID-19 et dont la méthodologie établie et largement documentée a facilité l'utilisation dans la production statistique. La volonté d'utiliser les données nouvelles et pleines de promesses des entreprises privées est de plus en plus forte, mais en l'absence d'incitations légales ou financières fortes, la situation est (et restera probablement) précaire.

## VIII. Conclusions et recommandations

59. Compte tenu des problèmes et enjeux évoqués dans la section précédente, les recommandations suivantes sont formulées aux niveaux international et national.

60. Au niveau international, il est proposé :

a) De renforcer la coopération, de recourir à la coordination stratégique dans l'élaboration des réglementations, des documents, des produits et des infrastructures informatiques clés, et de diffuser plus largement les résultats des travaux menés par les diverses équipes spéciales et les groupes de travail opérant au sein des principales organisations internationales ;

b) De spécialiser davantage les organisations internationales œuvrant à l'accès aux données détenues par le secteur privé, notamment par une meilleure répartition des tâches ou une focalisation sur des sources de données particulières. Une telle approche permettrait une utilisation plus efficace des ressources humaines et financières.

61. Au niveau national, il est proposé :

a) D'étudier et d'exploiter des vecteurs génériques de coopération avec les entreprises privées au niveau des experts, par exemple le détachement de fonctionnaires de l'organisme national de statistique pour œuvrer au sein des sociétés privées au traitement et à l'analyse des données ;

b) D'investir dans des partenariats, car la mise en place et la préservation d'accords de réutilisation des données exigeait une action soutenue ;

c) D'impliquer les pouvoirs publics ou les ministères concernés dans les discussions sur l'accès à des fins statistiques aux données détenues par le secteur privé et de solliciter leur appui dans les négociations avec les acteurs privés, si la législation sur les statistiques ne confère pas un poids suffisant à l'organisme national de statistique. À titre d'exemple, les opérateurs de réseaux mobiles pourraient être contactés avec l'aide du ministère compétent.

## IX. Délibérations du Bureau de la Conférence des statisticiens européens

62. Le Bureau a procédé à un examen approfondi de la collaboration avec les fournisseurs de données privés en février 2022 sur la base du document actuel (préparé par la Pologne (chef de file), le Canada, le Mexique, le Royaume-Uni, Eurostat et le FMI). Les observations suivantes ont été formulées au fil de la discussion :

a) Le document est parfaitement rédigé et propose une analyse intéressante des exemples de réussite et des enseignements tirés par les organismes nationaux de statistique et les organisations internationales en matière de coopération avec les fournisseurs de données du secteur privé ;

b) L'expression « données détenues par le secteur privé » couvre de nombreux types de données dotés de caractéristiques variées. Il serait utile de définir clairement ce que l'on entend par « données détenues par le secteur privé » ;

c) Il n'existe pas de base juridique solide permettant aux organismes nationaux de statistique de nouer le dialogue avec les fournisseurs de données privés et d'accéder à leurs données à des fins statistiques. La législation est un facteur essentiel pour offrir des garanties à l'appui de la confiance et des relations établies. L'accès doit être coordonné pour éviter que différentes agences gouvernementales ne contactent les mêmes entreprises privées ou ne paient pour les mêmes données ;

d) La collaboration avec les fournisseurs de données privés doit être profitable aux deux parties et sa pérennité suppose une bonne motivation. Il peut s'agir d'une référence à l'intérêt général, d'une rémunération financière, d'incitations législatives, de l'acceptation des consommateurs et du grand public, etc. ;

e) Les multinationales seront peut-être plus disposées à collaborer avec des organisations internationales qu'avec des organismes nationaux de statistique. Les organisations internationales peuvent jouer un rôle décisif et faciliter l'accès aux données détenues par le secteur privé. Les économies d'échelle sont également déterminantes ;

f) Une autre question importante est de savoir comment exploiter les données privées de manière sûre, sans enfreindre les règles de confidentialité. Cet aspect est à mettre en rapport avec la sensibilité de l'opinion publique et le débat sur la vie privée, très polarisé dans certains pays. Il est donc primordial d'établir une véritable entente avec la communauté de la protection de la vie privée ;

g) Les données détenues par le secteur privé sont sujettes à diverses erreurs d'échantillonnage et leur qualité n'est pas toujours évaluée comme il se devrait, ce qui complique leur intégration avec d'autres sources de données ;

h) Un document exprimant le point de vue commun des organismes officiels de statistique quant à l'accès aux données détenues par le secteur privé contribuerait à soutenir les efforts internationaux dans ce domaine. Il pourrait servir de vecteur de communication

efficace pour répondre aux préoccupations de l'opinion publique et du secteur privé (s'agissant notamment de la confiance et du respect de la vie privée) et établirait ainsi une sorte de « permis social » ;

i) Il conviendra de traduire le rapport sur l'examen approfondi dans d'autres langues officielles de la CEE et de le diffuser largement.

63. La session de la CSE du 22 juin 2022, consacrée à la collaboration avec les fournisseurs de données privés, sera l'occasion de clarifier la répartition des tâches dans ce domaine et les prochaines étapes requises à l'échelle nationale et internationale. Le Bureau analysera les résultats de la session et discutera de la nécessité d'un suivi en octobre 2022. Le Comité inter-États de statistique de la Communauté d'États indépendants a proposé de traduire le document en russe.

## X. Remerciements

64. L'Office polonais de statistique tient à remercier tout particulièrement Susie Fortier (Statistique Canada), Nicola Shearman (Office for National Statistics), Peter Struijs (Eurostat), Marco Marini (FMI), Paolo Quadrini (Institut national de statistique et de géographie), Jose Ruiz (Institut national de statistique et de géographie) et Andrea Fernandez (Institut national de statistique et de géographie) pour leur importante contribution à la préparation et à la réalisation de cet examen.

65. Nous remercions également tous les organismes nationaux de statistique et les organisations internationales qui ont répondu à l'enquête, d'avoir partagé leurs précieuses expériences et opinions.

## XI. Références

**Biancotti C., Borgogno O., Veronese G., 2021**, Principled data access: building public-private data partnerships for better official statistics, *Occasional Papers of Bank of Italy Series* n. 629.

**CSE 2019**, Nouvelles sources de données – accessibilité et utilisation, soixante-septième réunion plénière, point 2 a) de l'ordre du jour provisoire, Accéder à de nouvelles sources de 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 (directive vie privée et communications électroniques) (JO L 201, 31 juillet 2002, p. 37 à 47).

**ESSC 2020**, quarante-troisième réunion du Comité du système statistique européen à Luxembourg, point 4 de l'ordre du jour : Actions enabling the use of privately held data for official statistics, objectif n° 7 du programme de travail : Reaping the benefits of data revolution and moving to trusted smart statistics, 13 et 14 mai 2020.

**EU 2020**, Towards a European strategy on business-to-government data sharing for the public interest.

**OECD 2021**, Recommendation of the Council on Enhancing Access to and Sharing of Data, OECD/LEGAL/0463.

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (JO L 119, 4 mai 2016, p. 1 à 88).