

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

Genève, 22-24 juin 2020

Point 5 b) de l'ordre du jour provisoire

Nouveaux rôles des organismes nationaux de statistique et d'information géospatiale dans les récents écosystèmes nationaux de données :**Session 2 : Enseignements et résultats des mesures concrètes****déjà prises par les organismes nationaux de statistique et****les spécialistes de l'information géospatiale pour moderniser leur rôle****Géo-activation de la production statistique : de la conception à la diffusion¹****Note de Statistics Portugal***Résumé*

Le présent document décrit l'expérience de Statistics Portugal en matière de géo-activation de la production statistique. On y trouve un aperçu de projets et de produits élaborés à partir de données géospatiales ainsi que d'analyses et d'outils correspondants et mis en œuvre durant les différentes phases du modèle de production statistique, à savoir de la conception à la diffusion.

S'appuyant sur ces expériences, sur la participation de Statistics Portugal à des forums pan-européens et sur la production et l'utilisation de données géospatiales, le document présente aussi les principales difficultés rencontrées quant à l'intégration de données géospatiales dans la production de statistiques, ainsi qu'un ensemble de recommandations sur la façon de les surmonter.

Dans le cadre de la Conférence des statisticiens européens, le document est présenté pour examen au séminaire consacré au thème « Nouveaux rôles des organismes nationaux de statistique et d'information géospatiale dans les récents écosystèmes nationaux de données – Session 2 : Enseignements et résultats des mesures concrètes déjà prises par les organismes nationaux de statistique et les spécialistes de l'information géospatiale pour moderniser leur rôle ».

¹ Il a été convenu que le présent document serait publié après la date normale de publication en raison de circonstances indépendantes de la volonté du soumetteur.

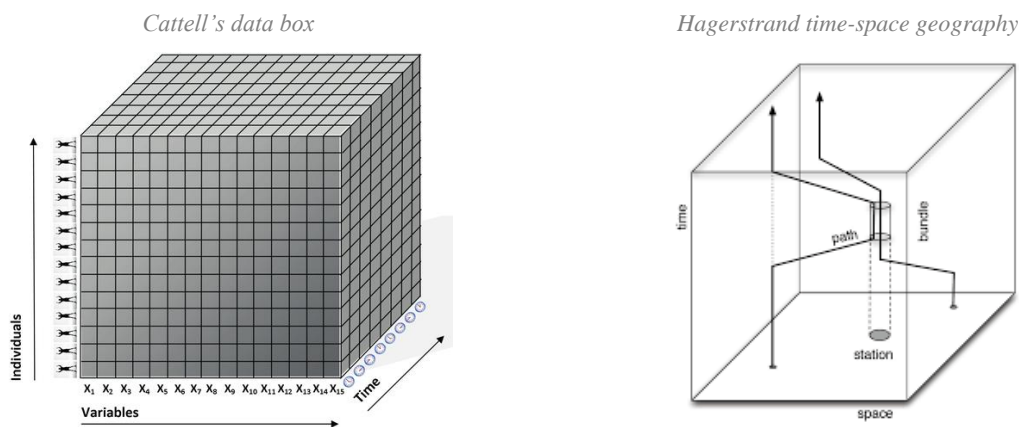


I. Introduction

1. Le paradigme de la production de données s'adapte à l'évolution de la société, à savoir la façon dont, au moyen des technologies, les individus, les organisations, voire les objets, interagissent entre eux en laissant des traces numériques de plus en plus nombreuses. Cette numérisation de la société signifie que de plus en plus les déplacements, les actions et les transactions sont enregistrés par un quelconque dispositif ou capteur numérique, ce qui permet de savoir CE QUI se passe, QUAND cela se passe mais aussi OÙ cela se passe.
2. L'*espace* (comme le temps) est un élément essentiel de la production statistique. Pour aborder cette dimension fondamentale des données, il est essentiel de recourir à des données géospatiales pour rendre correctement compte de l'élément « lieu » dans les différentes phases de la production statistique – de la conception à la collecte et à la gestion des données, puis à la diffusion – pour structurer et représenter les résultats statistiques et permettre une perception visuelle et territoriale des données.
3. Dans le modèle de production de données statistiques et avec le Modèle générique du processus de production statistique comme cadre de base, l'*espace* peut présenter trois dimensions importantes (Cordeiro *et al.*, 2012) : i) l'espace est fondamental pour *organiser* la collecte, le stockage, l'intégration et l'analyse de données ainsi que la diffusion des statistiques officielles ; ii) il devient important au niveau du *contexte* puisque les événements recueillis dans un segment territorial particulier varient selon les configurations territoriales appliquées pour décrire les résultats statistiques ; iii) il devient lui-même des informations statistiques car il *explique* et conditionne le phénomène considéré.
4. L'intégration de données est sur le point de passer d'un modèle de production statistique « en tuyau de poêle » à un modèle de production horizontal plus flexible qui favorise une réponse plus rapide et de meilleure qualité face aux nouvelles questions transversales, concernant notamment une plus grande granularité spatiale. Les informations géospatiales jouent un rôle majeur dans cette transformation de la production statistique en permettant un couplage plus précis des données et un appariement des données (spatiales) pour l'intégration de différents types de sources – des informations administratives publiques et privées aux mégadonnées et aux observations de la Terre. L'idée est de remplacer les modèles de données traditionnels, axés sur le projet statistique et sur une référence démographique particulière, par des modèles de données relationnels complexes qui intègrent divers domaines thématiques, sur la base de l'interaction d'agents centrée sur les activités menées dans l'espace et le temps [figure 1].

Figure 1

Modèle de données traditionnel d'un projet statistique et système théorique de relations d'agents dans le temps et l'espace



5. La combinaison des types de données, allant des sources de données traditionnelles telles que les enquêtes aux données administratives et, plus récemment, aux mégadonnées,

constitue une des dimensions clefs des écosystèmes de données. Il est essentiel d'ajouter de la valeur aux données en rassemblant les connaissances des statisticiens, des analystes géospatiaux et des scientifiques spécialistes des données pour créer un environnement de production statistique capable de repérer les changements intervenant dans la société et de fournir des données statistiques officielles pour les surveiller. En termes d'infrastructure, cela signifie qu'on travaille dans des environnements plus flexibles, numériques et adaptatifs, en utilisant de plus en plus les données et les logiciels en accès libre, les interfaces API, les systèmes de stockage en nuage, les centres de données et les plateformes partagées où les attributs de lieu sont essentiels. L'évolution vers une utilisation plus intensive et intégrée des données administratives et d'autres types de données est au cœur de la stratégie de Statistics Portugal, dont l'objectif est de créer une infrastructure de données nationale, dans laquelle les données géospatiales ainsi que les analyses et les outils correspondants jouent un rôle crucial.

6. L'intégration de données géospatiales dans le modèle de production de statistiques officielles a montré qu'elle augmentait la valeur des informations statistiques produites et diffusées. S'appuyant sur l'expérience acquise par Statistics Portugal en introduisant des données géospatiales dans les différentes phases du modèle de production statistique, le présent document a pour but de contribuer à l'examen des nouveaux rôles joués par les organismes statistiques et géospatiaux dans l'évolution vers une méthode de production intégrée. Sur la base de projets et de produits élaborés et diffusés par Statistics Portugal, le document présente les difficultés à surmonter et des recommandations visant à intégrer davantage des informations et des outils géospatiaux dans la chaîne de production statistique.

II. Intégration d'informations géospatiales dans la production de données statistiques

7. La géographie fait depuis longtemps partie de la production statistique, en particulier pour faciliter la préparation et la mise en œuvre de grandes opérations statistiques, comme le recensement de la population et des logements. Au Portugal, on peut faire remonter au début du siècle dernier l'utilisation de cartes pour promouvoir la diffusion des statistiques officielles. Plus récemment, il convient de mentionner l'utilisation de la cartographie associée au recensement de 2001. À cette époque, une « Base de référencement des informations géographiques » (BGRI 2001) a été créée à partir des systèmes d'information géographique (SIG). Pour le cycle de recensement de 2011, une base de référencement actualisée a été créée (BGRI 2011) et a été un outil important pour collecter, pour la première fois, les coordonnées x, y de tous les bâtiments recensés et pour établir une base de données par points. Ce type de données a contribué de manière notable à la production de la maille de population portugaise de 2011, également dans le cadre du projet GEOSTAT 2 du Système statistique européen (SSE) et de la diffusion de la maille de population européenne de 2011 – ensemble de données de maille GEOSTAT 2011 référencé par rapport au réseau de mailles INSPIRE de 1 km² (ETRS89-LAEA-1K).

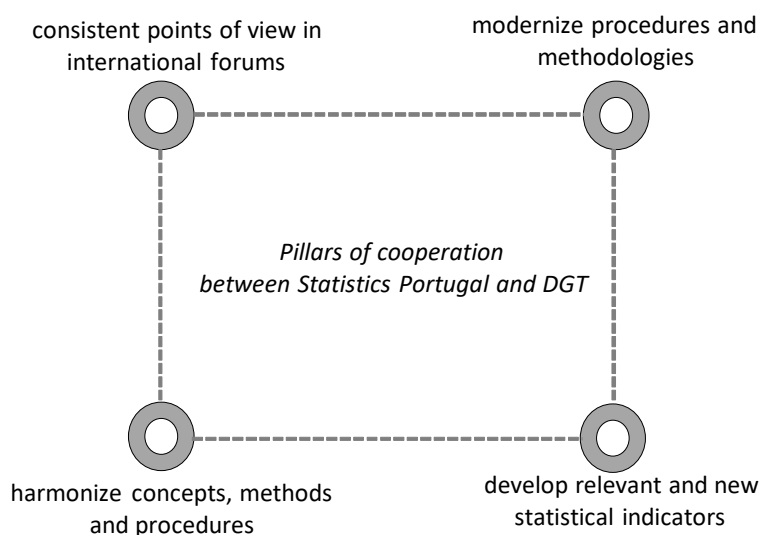
8. La directive INSPIRE (en vigueur depuis mai 2007) a également joué un rôle important dans l'harmonisation des données spatiales pour les thèmes de données pertinents et Statistics Portugal a participé aux travaux concernant 5 des 34 thèmes de la directive INSPIRE, à savoir : dénominations géographiques ; bâtiments et adresses – qui sont essentiels pour le registre des ménages et pour la mise en œuvre de davantage de processus de couplage et d'appariement de données ; unités statistiques ; répartition de la population et démographie. Au Portugal, la mise en œuvre de la directive INSPIRE est coordonnée par l'Administration cartographique et cadastrale nationale et la Direction générale du territoire.

9. Afin d'accroître l'interopérabilité entre les données géospatiales et les données statistiques, Statistics Portugal travaille en étroite collaboration avec les deux organismes susmentionnés et, en 2015, a établi un mémorandum d'accord qui prévoit quatre grands piliers de coopération, comme le montre la figure 2.

10. Outre qu'il contribue à élargir le champ d'intégration géographique et statistique dans la conception et la production d'indicateurs statistiques, le mémorandum d'accord offre également un cadre pour la modernisation et l'harmonisation des concepts et des méthodes, tout en gardant à l'esprit la nécessité de répondre aux normes de qualité de la production statistique.

Figure 2

Les quatre piliers de la coopération entre Statistics Portugal et la Direction générale du territoire



11. L'une des instances internationales auxquelles Statistics Portugal participe activement, en collaboration avec l'Administration cartographique et cadastrale nationale (Direction générale du territoire), est le Groupe de travail de l'intégration des données du Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale. Le Groupe de travail s'est consacré à la géo-activation des indicateurs du développement durable et a publié en mai 2019, sous la houlette de Statistics Portugal, un rapport intitulé *The territorial dimension in SDG indicators: Geospatial data analysis and its integration with statistical data* (Dimension territoriale dans les indicateurs des objectifs de développement durable (ODD) : analyse des données géospatiales et intégration aux données statistiques). L'un des principaux résultats statistiques des travaux réalisés dans le cadre de ce rapport a été le calcul et la diffusion d'une *information indirecte* sur l'indicateur 11.3.1 des ODD, à savoir le *ratio entre taux d'utilisation des terres et le taux de croissance démographique*, fondé sur la cartographie de l'utilisation des terres et du couvert terrestre produite par l'Administration cartographique et cadastrale nationale (Direction générale du territoire). Dans le cadre du plan de travail 2019-2022 du Groupe de travail, Statistics Portugal continuera à diriger les tâches consacrées à la géo-activation des indicateurs ODD, en privilégiant les indicateurs liés à l'environnement et l'utilisation des données provenant de l'observation de la Terre.

12. L'utilisation des données de télédétection à des fins statistiques date de longtemps, en particulier pour les statistiques agricoles (CEE, 2019). En 2015, dans le cadre du mémorandum d'accord, Statistics Portugal et la Direction générale du territoire ont mené une étude pilote (subvention du SSE²) pour étudier les données de télédétection et des sources de données nationales supplémentaires afin de produire des statistiques sur l'occupation des sols au niveau NUTS 3, en remplacement du système LUCAS, qui s'appuie sur des données *in situ* collectées par les enquêteurs (Costa *et al.*, 2018). Actuellement, Statistics Portugal participe aussi à un ensemble de travaux sur les mégadonnées relatives à l'observation de la Terre dans le cadre du réseau ESSnet. Ces travaux concernent principalement les données satellitaires et les photographies aériennes

² EUROSTAT – Contrat n° 08441.2015.002-2015.724 – Fourniture d'informations harmonisées sur l'occupation et l'utilisation des sols : système LUCAS et systèmes nationaux.

(données Sentinel), le but étant de définir un cadre géospatial pour la ventilation des données entre les informations statistiques et géographiques. L'accent est mis sur la disponibilité des données et les conditions d'accès pertinentes pour les domaines statistiques, tels que l'agriculture, la sylviculture ou le dénombrement des établissements.

13. Tirant parti de l'utilisation accrue et diversifiée de la technologie SIG au sein des organismes de statistique, Statistics Portugal axe sa stratégie à moyen terme sur la nécessité de promouvoir une plus grande interopérabilité entre les données spatiales et les données statistiques afin d'appuyer la production statistique et de favoriser l'intégration spatiale et statistique. L'objectif est de produire de nouveaux indicateurs statistiques, dans un souci permanent d'introduire une perspective spatiale dans les différentes phases de la production statistique, comme le montrent ci-après les projets et produits mis au point dans les différentes phases.

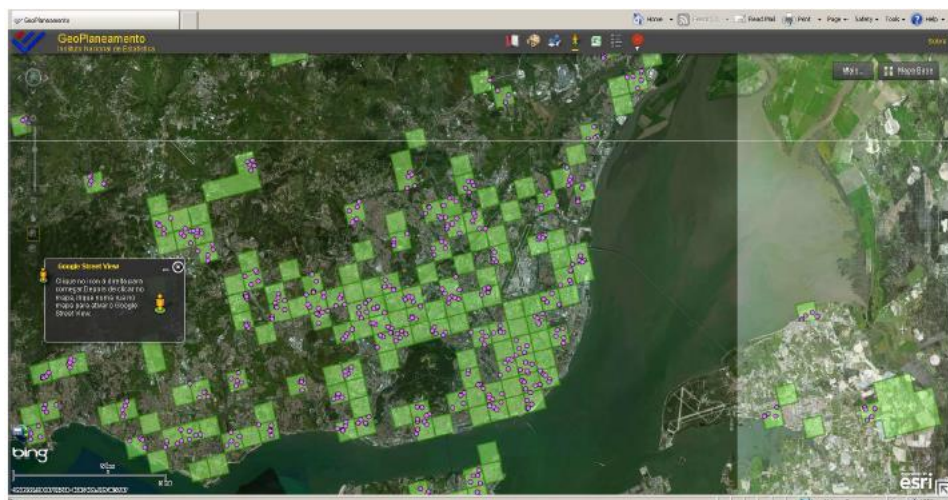
A. Utilisation d'un modèle d'échantillonnage spatial

14. Dans le cadre d'une stratégie visant à accroître la modernisation et l'efficacité de la production statistique et en apportant des améliorations méthodologiques et technologiques, Statistics Portugal a mis en pratique une nouvelle méthode pour définir le modèle et les bases d'échantillonnage. En s'appuyant sur les informations géoréférencées (coordonnées x, y) de tous les bâtiments recensés en 2011, l'organisme a défini un fichier national des logements pour appuyer le processus d'échantillonnage des enquêtes auprès des ménages, régulièrement mis à jour par des données administratives. Un important instrument géospatial a également été intégré dans ce processus, la maille européenne de 1 km² (réseau de mailles ETRS89-LAEA-1K de l'infrastructure INSPIRE) comme nouvelle référence pour le choix de l'unité d'échantillonnage primaire.

15. Habituellement, le choix s'opère selon un schéma d'échantillonnage stratifié et à plusieurs degrés, dans lequel les unités d'échantillonnage primaires, géographiquement constituées d'une ou de plusieurs cellules contiguës de 1 km² [figure 3], sont systématiquement choisies avec une probabilité proportionnelle à l'importance du nombre de lieux de résidence habituelle (logements) ; les unités d'échantillonnage secondaires sont systématiquement choisies parmi les unités retenues lors de la première étape. Toutes les unités d'échantillonnage primaires des bases d'échantillonnage des enquêtes comportant des rotations doivent inclure les routes.

Figure 3

Exemple de choix de mailles en vue de la définition des unités d'échantillonnage primaires



16. L'utilisation de ce modèle d'échantillonnage spatial, associé au choix des logements dans les « segments », a permis de réduire le coefficient de corrélation intra-grappes (qui mesure la similarité des unités statistiques).

17. Il a été démontré qu'une base d'échantillonnage géoréférencée améliore la précision des estimations. Plus le modèle d'échantillonnage choisit des individus géographiquement éloignés les uns des autres, plus l'estimation sera précise pour une variable spatialement auto-corrélée (Favre-Martinoz *et al.*, 2018). En outre, dans le cas d'entretiens en face à face, en connaissant l'emplacement des unités statistiques échantillonnées, on peut plus aisément les recenser sur le terrain et gérer le lieu où se trouvent les enquêteurs lorsqu'ils travaillent sur le terrain. Il est primordial d'actualiser en permanence les données ponctuelles sous-jacentes pour accroître l'efficacité du processus de conception des échantillons spatiaux et de la collecte de données.

B. Renforcement de l'efficacité de la gestion de la collecte de données grâce aux outils géospatiaux

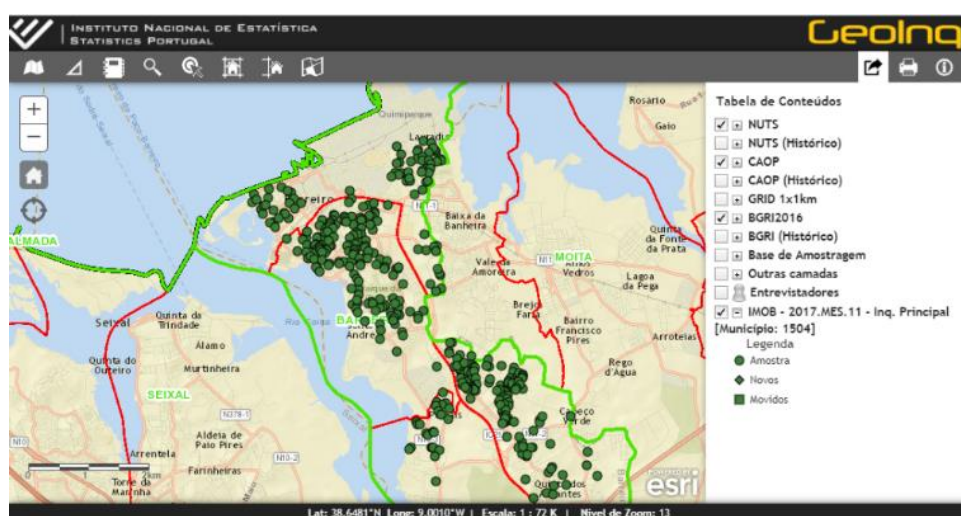
18. La collecte de données est un domaine clef de la production statistique. Un objectif important est de développer des procédures et des outils qui permettent aux répondants de fournir plus facilement des informations, tout en essayant de réduire leur charge de travail. Cela étant, une dimension fondamentale du processus consiste aussi à rechercher des solutions qui permettent aux enquêteurs d'agir de la manière la plus productive possible pour accroître l'efficacité de la collecte de données.

19. À Statistics Portugal, les enquêteurs ont régulièrement eu du mal à localiser les logements échantillonnés dans les enquêtes auprès des ménages, car ils ne pouvaient se fier qu'aux tableaux contenant les coordonnées, le nom et les contacts du représentant du ménage. Un outil géospatial en ligne, conçu sur mesure pour répondre aux besoins en matière de production de données statistiques, a été mis en place, dans le cadre de l'intégration des données géospatiales dans le modèle de production de statistiques officielles.

20. L'application en ligne GeoINQ [figure 4] a été développée par Statistics Portugal en partenariat avec la société ESRI à l'aide d'une interface API destinée à l'environnement ArcGIS. L'outil intègre des données ponctuelles concernant les ménages figurant dans les bases d'échantillonnage et un ensemble de couches géospatiales de fond pertinentes (nomenclature NUTS, unités administratives, grille de 1 km², base de référencement des informations géographiques) et des cartes de base, y compris les orthophotocartes de l'Administration cartographique et cadastrale nationale (Direction générale du territoire).

Figure 4

Application en ligne GeoINQ



21. Avec l'application GeoINQ, les enquêteurs peuvent facilement déterminer la localisation précise (x, y) des logements et avoir accès aux données connexes. Cette application fonctionne sur les appareils mobiles et les utilisateurs ne peuvent accéder qu'aux caractéristiques et aux couches géographiques compatibles avec leur profil d'utilisateur.

22. L'application est entièrement intégrée à d'autres systèmes mis au point par Statistics Portugal, en particulier au système mondial de gestion des enquêtes par entretiens (SIGINQ-IE). Par conséquent, outre les enquêteurs, d'autres utilisateurs internes ont recours à cette application en ligne pour répondre à leurs besoins en matière de gestion et d'analyse des données, notamment pour analyser la dispersion géographique et le chevauchement des échantillons sur le territoire national dans le cadre du processus de conception de l'échantillonnage spatial, comme le décrit la section précédente. Le but recherché est aussi d'aider et de gérer les enquêteurs dans leur travail sur le terrain, y compris par l'attribution des échantillons. Il est fondamental, dans ce contexte, d'actualiser en permanence les données géospatiales sous-jacentes pour continuer à tirer profit des fonctionnalités utiles liées à ce type d'outil géospatial qui facilite la production de données statistiques.

C. Mise en œuvre de solutions géospatiales pour recueillir les variables complexes

23. En 2017, Statistics Portugal a mené une enquête sur la mobilité dans les deux zones métropolitaines de Porto et de Lisbonne. En s'appuyant sur un échantillon aléatoire stratifié et multiphasique, qui a pris en compte des zones homogènes en matière d'accessibilité des transports, l'organisme a appliqué une méthode mixte de collecte de données en associant les entretiens en ligne assistés par ordinateur et les entretiens individuels assistés par ordinateur.

24. L'enquête avait pour objectif de caractériser les déplacements (non limités aux migrations pendulaires) de la population résidente (âgée de 6 à 84 ans) dans les deux zones métropolitaines, ce qui supposait la capacité de capturer les points d'origine et de destination de chaque déplacement pendant un jour précis de la semaine, ainsi que d'autres dimensions afin de comprendre comment les individus se déplacent, à quelle fréquence ils voyagent, combien de temps ils passent à se déplacer, où ils vont et pour quoi faire. La principale difficulté liée à la conception d'une enquête en ligne visant à atteindre cet objectif consistait à trouver un moyen de permettre aux gens d'enregistrer facilement leurs déplacements pendant la journée et de trouver et de déterminer avec précision les endroits où ils se rendent.

25. À la place de rapports descriptifs, l'organisme a appliqué une solution novatrice en recourant à Google Maps. Les cartes ont servi à capturer les destinations des déplacements au moyen des mêmes fonctions que celles utilisées par les gens dans Google Maps et au moyen de cercles de localisation basés sur le centroïde de la municipalité allant jusqu'au point le plus éloigné, aidant ainsi les gens à naviguer dans les différents lieux [voir la figure 5].

Figure 5

Exemple d'écran de réponse pour la détermination des lieux dans l'enquête sur la mobilité dans les zones métropolitaines

26. L'externalisation de services à des fins statistiques n'est cependant pas exempte d'une évaluation de leurs hypothèses de base effectuée pour s'assurer qu'ils répondent aux critères de qualité de la production statistique. Cette évaluation peut être plus limitée pour les bases de données et les produits commerciaux. Cela suppose également une dépendance à l'égard des services extérieurs assortie d'une capacité d'intervention limitée et une exposition à des changements qui peuvent affecter directement ou indirectement les processus de production statistique mis en œuvre.

D. Production d'indicateurs statistiques pour le suivi des objectifs de développement durable au niveau territorial

27. Récemment, l'adoption du Programme de développement durable à l'horizon 2030 (Nations Unies, 2015) et la définition de 17 objectifs de développement durable (ODD) devant être suivis par 232 indicateurs ont souligné l'importance de la désagrégation géographique des données (concernant, par exemple, les zones urbaines par rapport aux zones rurales) ainsi que d'autres segmentations lorsqu'il s'agit d'observer la devise « *ne laisser personne derrière* ». Au niveau européen, un ensemble d'indicateurs a été mis en place pour mesurer les progrès réalisés en matière d'ODD dans un contexte européen (Eurostat, 2019). Statistics Portugal a rassemblé les informations disponibles pour le Portugal selon le cadre de suivi mondial des ODD³. Depuis 2018, un rapport annuel est également publié (par exemple, INE, 2019). Il comporte une brève analyse des résultats de chaque indicateur disponible (de 2010 à l'année la plus récente), y compris des données géographiquement ventilées aux niveaux régional (NUTS 2 et 3) et municipal.

28. Dans le cas du Portugal, la couverture des indicateurs statistiques est plus faible pour certains ODD, surtout si le cadre de suivi comprend des indicateurs des niveaux II et III⁴. Par conséquent, Statistics Portugal s'est attelé à la tâche pertinente qu'est l'élargissement de la portée des indicateurs ODD disponibles, en particulier au niveau territorial. Plus

³ Une [section consacrée aux ODD](#) a été publiée sur le site Web de Statistics Portugal.

⁴ Au niveau mondial, les indicateurs ont été classés selon un système à trois niveaux concernant la disponibilité des données et la méthode établie ; i) pour les indicateurs de niveau I, une méthode a été établie et des données sont déjà largement disponibles ; ii) pour les indicateurs de niveau II, une méthode a été établie, mais les données ne sont pas facilement disponibles ; iii) pour les indicateurs de niveau III, aucune méthode n'a encore été établie et approuvée à l'échelle internationale.

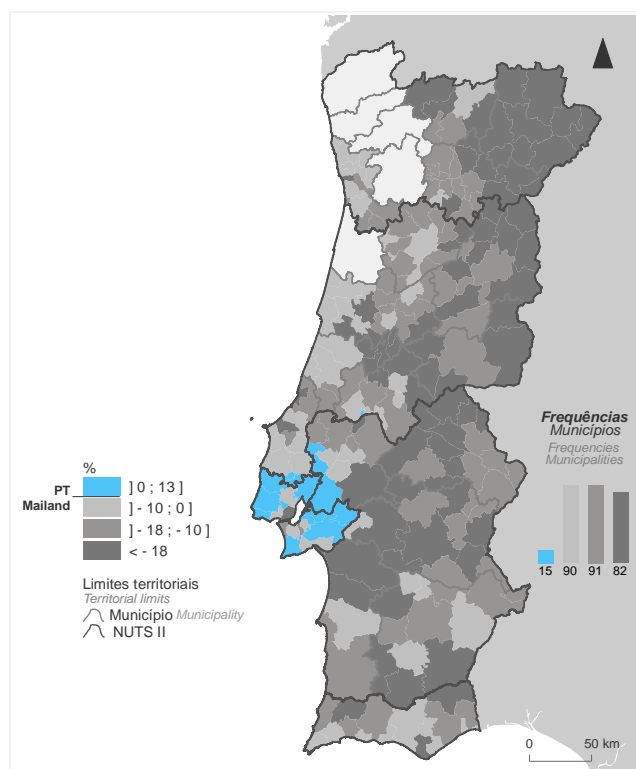
précisément, l'intégration des données géospatiales et statistiques et l'analyse géospatiale ont permis de réaliser des progrès pour ce qui d'accroître la portée des informations nécessaires au suivi de l'objectif 11 concernant les villes et les communautés durables.

29. En 2018, Statistics Portugal a publié une nouvelle série de statistiques sur l'occupation des sols et du couvert végétal (LCLUStats) fondées sur la carte de l'occupation des sols et du couvert végétal produite par l'Administration cartographique et cadastrale nationale (Direction générale du territoire) à partir de la photo-interprétation d'images aériennes orthorectifiées. Ces statistiques comportent le calcul, au niveau des municipalités, d'une *information indirecte* sur l'indicateur de niveau II de l'ODD 11.3.1 (*ratio entre le taux d'utilisation des terres et le taux de croissance démographique*) fondé sur la formule de l'efficacité de l'utilisation des terres (Corbane *et al.*, 2017), telle que proposée par le Centre commun de recherche (CCR). Cette formule combine les données de la carte de l'occupation des sols et du couvert végétal et les estimations annuelles de la population résidente pour les années de référence de la carte (2010 et 2015). Les résultats sont normalisés pour une période de dix ans.

30. Le résultat pour le Portugal continental était de -10 % pour la période 2010-2015. Seules 15 municipalités, principalement situées dans la zone métropolitaine de Lisbonne, ont obtenu des valeurs positives en matière d'efficacité de l'utilisation des terres, ce qui signifie que la croissance démographique est plus rapide que l'augmentation des terres artificielles. Pour un groupe de 90 municipalités, situées principalement dans la zone côtière de la région septentrionale et de la région centrale, l'efficacité de l'utilisation des terres a baissé, mais cette baisse est encore moins importante que la valeur moyenne applicable au Portugal continental (-10 %) [figure 6].

Figure 6

Efficacité de l'utilisation des terres par municipalité (2015)



31. Comme il s'agit de la première opération statistique diffusée par Statistics Portugal à partir d'une source de données géospatiales et de son intégration à des données statistiques, la diffusion a posé quelques problèmes lors de la prise en compte des données géospatiales et de la réalisation de l'analyse selon le document méthodologique statistique normalisé, qui décrit la totalité des procédures, des concepts et des classifications associés à une opération statistique.

E. Utilisation d'outils géospatiaux en accès libre pour mesurer l'accessibilité des services

32. L'accessibilité des services est une dimension pertinente pour la mesure du bien-être et de la qualité de vie, qui sont devenus des aspects importants de l'évaluation réalisée par les décideurs pour mieux appréhender les progrès accomplis au niveau de la société et des conditions de vie (voir, par exemple, l'initiative *How's life* (Comment va la vie ?) de l'OCDE). Le Programme de développement durable à l'horizon 2030 des Nations Unies souligne également que l'accessibilité est une dimension pertinente pour le suivi de l'objectif 11 concernant les villes et les communautés durables et a inclus un indicateur pour le suivi de l'accessibilité des transports publics. Cet indicateur a cependant été défini comme étant de niveau II, ce qui signifie qu'une méthode a été établie pour le calcul, mais que les données ne sont pas facilement disponibles.

33. Dans le cadre d'une subvention du Système statistique européen (SSE) concernant les statistiques infranationales (Audit urbain, 2017-2019)⁵, Statistics Portugal a mis au point une activité visant à accroître les connaissances sur la mesure des indicateurs d'accessibilité. Il s'agissait essentiellement d'étudier l'accessibilité des écoles. Des mesures expérimentales de la couverture territoriale et démographique ont été calculées compte tenu des distances à pied et en voiture à partir de l'école, les isochrones étant définies entre 5 et 40 minutes avec des intervalles de temps de 5 minutes [figure 7]. Ces zones de service ont été calculées à l'aide de données et de logiciels en accès libre, à savoir le réseau de navigation *Open Street Map* (OSM) grâce au plug-in *Open Route Service* (ORS) dans l'environnement SIG Quantum. La proportion de territoire (superficie) et de population (données ponctuelles du recensement de 2011) couverte par les écoles a été calculée pour différentes unités territoriales, y compris au niveau de la maille [figure 8] et de la municipalité [figure 9].

⁵ EUROSTAT/Contrat n° : 08142.2017.002-2017.432 – Collecte de données pour les statistiques infranationales (principalement les villes).

Figure 7
Zones de service des établissements
d'enseignement de base situées à 5 à
40 minutes à pied

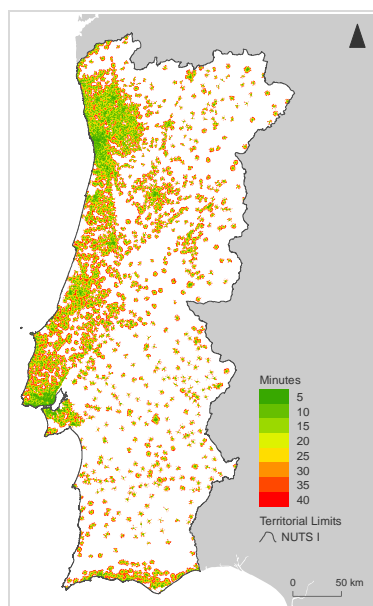


Figure 8
Couverture démographique des
établissements d'enseignement de
base situés à 15 minutes à pied par
maille de 1 km²

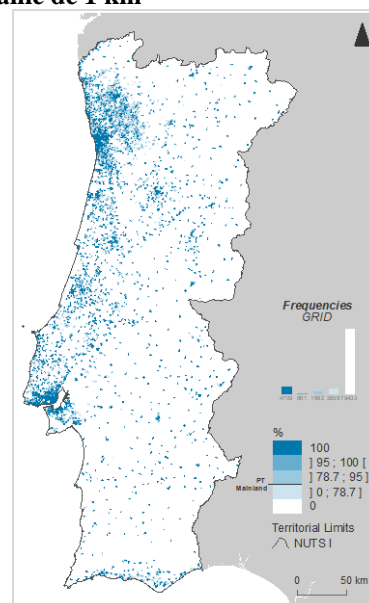
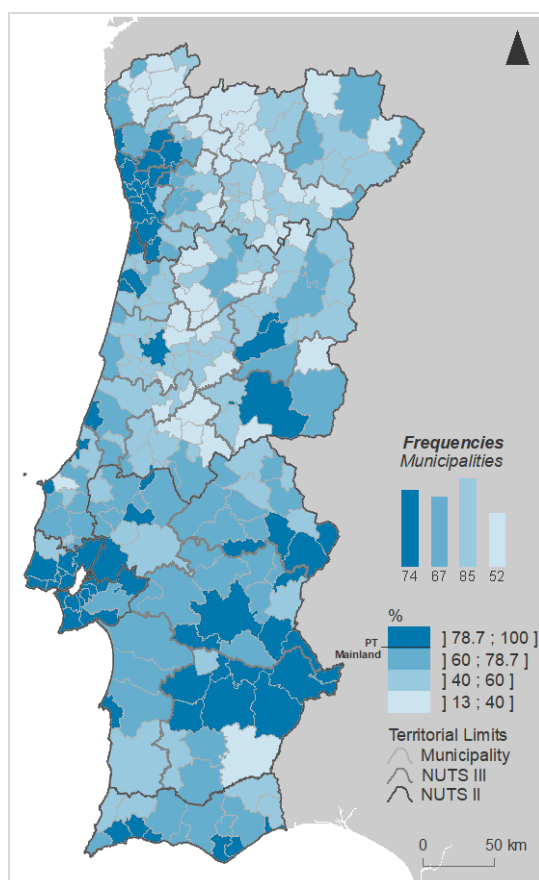


Figure 9

Couverture démographique des établissements d'enseignement de base situés à 15 minutes à pied par municipalité



34. Étant donné le caractère expérimental de ces indicateurs d'accessibilité, il a été procédé à une analyse comparative afin de déterminer la qualité des données. Pour les mêmes origines et destinations, certains résultats ont été comparés à ceux d'autres solutions disponibles et il a pu être constaté que les données relatives aux distances parcourues à pied semblent être plus fiables que celles concernant les distances parcourues en voiture. Par conséquent, et bien que l'utilisation de données et d'outils SIG en accès libre ait permis de pallier l'absence d'un réseau de navigation officiel actualisé dans le contexte du Portugal, il est important de comparer les résultats obtenus à d'autres sources pour évaluer la cohérence et la fiabilité des résultats et ainsi produire des indicateurs statistiques officiels de l'accessibilité.

F. Création d'outils de visualisation des données géospatiales

35. Suite à la crise financière et économique internationale, il est devenu de plus en plus urgent de disposer d'informations territoriales sur les prix des logements pour suivre les changements intervenus sur le marché des logements au Portugal. À l'échelle de l'Union européenne (UE), en collaboration avec les États membres, Eurostat a également développé des outils statistiques permettant d'analyser l'évolution du marché immobilier, notamment celui des logements (Eurostat, 2018).

36. En 2017, Statistics Portugal a commencé à diffuser des statistiques trimestrielles sur les prix des logements à l'échelon local en se fondant sur des données administratives fiscales géoréférencées, à savoir la taxe municipale sur les transferts de propriété (qui permet d'obtenir les prix des transactions) et la taxe foncière municipale (qui permet d'obtenir les caractéristiques d'un logement faisant l'objet d'une transaction, y compris les



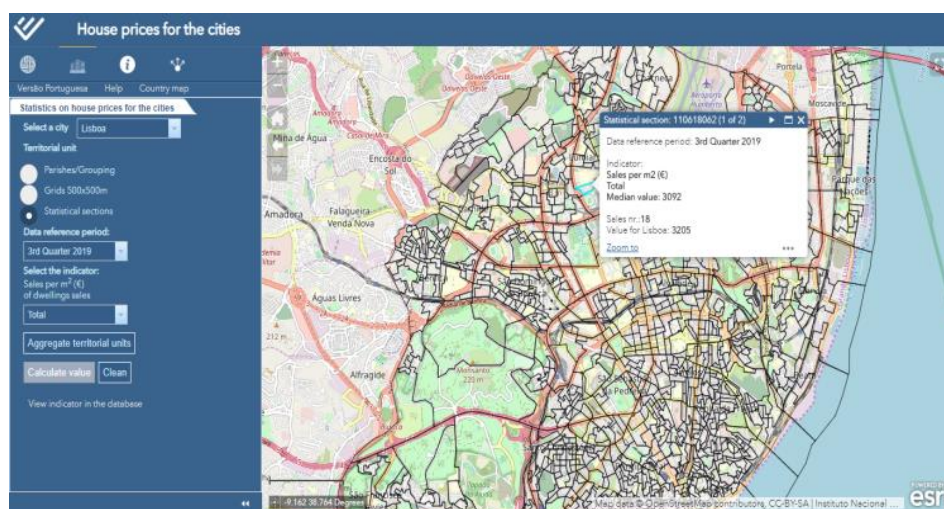
coordonnées x, y et les plus petites unités administratives locales – les paroisses), ces données ayant été fournies par l'Administration des impôts et des douanes.

37. Outre la diffusion régulière d'indicateurs statistiques selon les unités territoriales communes (niveau des municipalités et des paroisses de la nomenclature NUTS), Statistics Portugal avait également pour objectif de fournir un outil qui permettrait aux utilisateurs de parcourir les informations à un niveau géographique local plus détaillé. La solution a été trouvée grâce à un outil de visualisation des données géospatiales qui offre aux utilisateurs la possibilité de personnaliser la recherche des prix des logements en fonction de différentes zones géographiques. Les données de l'application en ligne ne comprennent que les registres dont les coordonnées x, y sont valides, après l'application d'une procédure de validation et un couplage des informations complémentaires de l'Agence portugaise de l'énergie avec les données IMT et IMI (à l'aide de la variable « Code de logement de l'Administration fiscale »). Les coordonnées géographiques et le codage des unités administratives locales sont également validés sur la base de la carte administrative officielle du Portugal.

38. L'outil « [Prix des logements – Villes](#) », qui a été conçu à l'aide d'une interface API (JavaScript) pour l'environnement ArcGIS, est compatible avec les appareils mobiles. Cette application en ligne permet de rechercher les prix de vente médians des logements (€/m²) dans les sept villes portugaises de plus de 100 000 habitants – Lisbonne, Porto, Vila Nova de Gaia, Amadora, Braga, Funchal et Coimbra [figure 10].

Figure 10

Application en ligne concernant les prix des logements dans les villes – Lisbonne



Source : Statistics Portugal, Statistiques sur les prix des logements à l'échelon local

39. Les utilisateurs peuvent parcourir et personnaliser les données choisies par paroisse, par section statistique (zones géographique du recensement de 2011) et par maille de 500 m x 500 m. Pour les sections statistiques et les mailles, les résultats se rapportent à sept transactions au minimum.

40. L'application en ligne concernant les prix des logements est l'un des produits les plus consultés de Statistics Portugal, ce qui témoigne de son utilité et de sa capacité à répondre aux besoins des utilisateurs. Étant donné l'intérêt que présentent les coordonnées x, y pour la présentation en tableaux des données au niveau des villes, il est essentiel de mettre en œuvre des procédures de validation, portant notamment sur la cohérence avec les unités des divisions administratives et l'utilisation de sources de données auxiliaires (Agence nationale pour l'énergie) pour accroître la disponibilité des données et assurer la qualité de celles-ci. De plus, le niveau de granularité des données signifie qu'il faut soigneusement en évaluer la fiabilité. La médiane a été prise comme paramètre de référence pour la diffusion des prix des logements à l'échelle locale, le but étant de mieux remédier à la forte asymétrie des répartitions et aux problèmes de confidentialité soulevés par la possibilité de sélectionner des données personnalisées selon les configurations territoriales définies par les utilisateurs.

III. Difficultés et recommandations

41. Les exemples précédents et l'expérience globale de Statistics Portugal en matière d'intégration d'informations géospatiales dans la production de données statistiques permettent d'exposer ci-après un certain nombre de difficultés liées aux principes statistiques définis par le Code de bonnes pratiques de la statistique européenne⁶, qui pourraient également être d'actualité pour d'autres pays et dans le cadre européen général. Sont présentées ensuite les recommandations à suivre pour venir à bout de ces difficultés.

A. Respect du principe statistique « Engagement sur la qualité » (principe 4)

42. L'utilisation des données géospatiales ainsi que des analyses et des outils correspondants, tout comme l'intégration de ces données à des données statistiques, a permis d'obtenir de nouvelles informations pertinentes pour traiter des questions transversales et répondre aux défis mondiaux, comme c'est le cas du cadre de suivi des ODD. La qualité des données doit néanmoins être assurée lorsqu'on utilise des sources de données et des outils non officiels, qu'ils soient de nature commerciale (comme dans l'exemple de l'utilisation d'une interface API Google Maps pour recueillir des données sur les lieux et calculer les distances aux fins de l'Enquête sur la mobilité) ou en accès libre (comme dans le cas de l'utilisation des données OSM (*Open Street Map*) et de l'outil ORS (*Open Route Service*) dans l'environnement QGIS pour calculer les indicateurs d'accessibilité des écoles). Dans ce contexte, il est essentiel de vérifier la stabilité, la cohérence et la fiabilité des données en établissant un rapport complet sur les métadonnées et en comparant les résultats. L'utilisation de ces outils analytiques et sources de données géospatiales souligne en outre l'intérêt qu'il y a à disposer de données et d'outils géospatiaux officiels bien étayés et certifiés pour produire des résultats statistiques.

B. Respect du principe statistique « Méthodologie saine » (principe 7)

43. L'éventail des informations géospatiales du ressort des activités statistiques ne se limite pas aux données géospatiales collectées par Statistics Portugal ou par l'Administration cartographique et cadastrale nationale (Direction générale du territoire). Plusieurs autres entités de l'administration publique produisent des données géospatiales pertinentes dans le cadre de leurs activités. Cela étant, différentes approches méthodologiques entrent en jeu, ce qui entrave et compromet la compatibilité et l'interopérabilité des données. C'est le cas, par exemple, des données ponctuelles utilisées pour connaître les *prix des logements à l'échelle locale* (fondées sur les données de l'Administration fiscale portugaise) et des données géoréférencées issues du recensement de 2011 concernant les bâtiments (produites par Statistics Portugal) qui ne sont pas compatibles ni avec les systèmes de codage ni avec les normes de géoréférencement. La coordination à cet égard est donc essentielle aux niveaux national et européen.

C. Respect du principe statistique « Secret statistique et protection des données » (principe 11)

44. Un défi pour le maintien du secret des données réside dans l'accroissement de la granularité des données ainsi que dans la production et la diffusion de données selon des zones géographiques de niveau très détaillé, y compris la possibilité de choisir des configurations territoriales particulières, comme dans le cas de l'utilisation d'un outil de

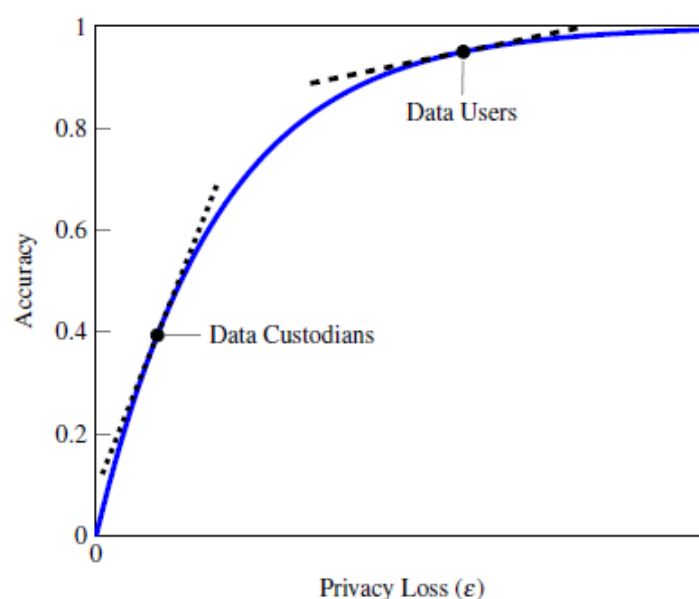
⁶ Le Code (version révisée de 2017) « contient 16 principes liés à l'environnement institutionnel, aux processus de statistiques et à la production statistique. Le Code a pour but d'assurer que les statistiques produites dans le cadre du Système Statistique Européen (SSE) sont pertinentes, disponibles à temps et précises et qu'elles respectent les principes d'indépendance professionnelle, d'impartialité et d'objectivité ». (<https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>).

visualisation géospatial pour la diffusion des statistiques sur les prix des logements à l'échelle locale (Prix des logements – Villes). Il est nécessaire de mettre en pratique une évaluation critique pour garantir la protection des données, tout en essayant de répondre aux besoins des utilisateurs [figure 11].

45. Par ailleurs, le fait d'enrichir le contenu géospatial dans la production de données statistiques peut avoir une incidence positive sur les méthodes et les procédures de contrôle de la divulgation des données statistiques visant à garantir le secret.

Figure 11

Rapport coût-bénéfice entre l'atteinte à la vie privée et le détail des informations



D. Respect du principe statistique « Engagement sur la qualité » (principe 14)

46. La disponibilité de sources de données géospatiales nationales et leur intégration à des données statistiques permettent aux pays de disposer d'indicateurs statistiques et de typologies nationales assortis d'une ventilation territoriale plus élevée qui sont utiles à la formulation et au suivi des politiques fondées sur le territoire. Cela peut cependant entraîner des différences conceptuelles et méthodologiques par rapport au cadre réglementaire établi par le Système statistique européen (SSE) pour un domaine précis, ce qui peut nuire dans certains cas, à la comparabilité avec d'autres pays. Par exemple, les statistiques nationales sur l'occupation des sols et du couvert végétal (LCLUStats) fournissent des données détaillées pertinentes, notamment jusqu'au niveau des municipalités, sur la situation et les changements observés dans ce domaine et éclairent ainsi les politiques nationales d'aménagement régional. Celles-ci sont établies à partir de la carte nationale de l'occupation des sols et du couvert végétal, qui repose sur une méthode différente de celle utilisée par l'UE dans l'Enquête LUCAS pour fournir des statistiques harmonisées et comparables sur l'occupation des sols et du couvert végétal dans les régions de l'UE, mais seulement jusqu'au niveau NUTS 2.

Recommandations visant à géo-activer la production statistique⁷

Harmoniser les thèmes communs relatifs aux données géospatiales à l'échelle européenne, en tenant compte des caractéristiques des données de base pour l'analyse spatiale et l'intégration des données à des fins statistiques (par exemple, métadonnées, échelles, attributs, précision) et en suivant les recommandations ayant trait aux données de base du Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale, en complétant les spécifications de la directive INSPIRE sur les données et en définissant les priorités en matière de contenu de base pour répondre aux besoins des utilisateurs et traiter les ODD.

Mettre en œuvre les thèmes communs relatifs aux données géospatiales clefs, comme les bâtiments, les adresses, l'occupation des sols et du couvert végétal, les données cadastrales, les réseaux de transport, en tant que données faisant autorité à l'échelle européenne, l'Administration cartographique et cadastrale nationale assumant un rôle de coordination pertinent au niveau national.

Assurer la disponibilité des sources de données géospatiales et des outils de traitement, d'analyse et de visualisation à l'échelle européenne et l'accès à ces sources pour permettre la production de statistiques géospatiales de manière harmonisée et cohérente dans les États membres.

Accroître l'harmonisation et l'interopérabilité des données géospatiales produites par les organismes nationaux dans le cadre de la définition et de la mise en œuvre d'une stratégie nationale en matière de données spatiales, en tenant compte des exigences de la production de données statistiques.

Élargir la communication et les liens entre les producteurs de données géospatiales, les organismes de statistique ainsi que les scientifiques et les chercheurs spécialistes des données pour tirer parti de l'infrastructure nationale des données spatiales et de l'intégration des données géospatiales et des données statistiques.

⁷ Les présentes recommandations sont formulées à partir des débats menés au sein du Groupe de travail de l'intégration des données relevant du Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale, et plus particulièrement à partir des résultats obtenus (Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale, 2019a et 2019b).

IV. Bibliographie

- Abowd, J. M., Schumtte, I. M., Sexton, W. & Vilhuber, L. (2019) Why the economics profession must actively participate in the privacy protection debate. *American Economic Association Papers and Proceedings*, 109: 397-402.
- Corbane, C., Politis, P., Siragusa, A., Kemper, T. & Pesaresi, M. (2017). LUE User Guide: A tool to calculate the Land Use Efficiency and the SDG 11.3 indicator with the Global Human Settlement Layer. Luxembourg : Office des publications de l'Union européenne.
- Cordeiro, H., Vala, F. & Santos, A. (2012). Spatial Data Infrastructure for statistical production: challenges and opportunities. Paper presented at 98th Directors General of the National Statistical Institutes (DGINS) Conference, Prague, 24-25 septembre.
- Costa, H., Almeida, D., Vala, F., Marcelino, F., & Caetano, M. (2018). Land cover mapping from remotely sensed and auxiliary data for harmonized official statistics. *ISPRS International Journal of Geo-Information*, 7(4), 1-21. [157]. DOI: 10.3390/ijgi7040157.
- Eurostat (2019). Sustainable development in the European Union: Monitoring report on the progress towards the SDGs on an EU context. Luxembourg : Office des publications de l'Union européenne.
- Eurostat (2018). Housing price statistics - house price index. Statistics explained. Disponible à l'adresse http://ec.europa.eu/eurostat/statisticsexplained/index.php/Housing_price_statistics_-_house_price_index.
- Favre-Martinoz, C., Fontaine, M., Le Gleut, R. & Loonis, V. (2018). Spatial sampling. In V. Loonis and M.P Bellefon, *Handbook of Spatial Analysis: Theory and Application with R* (pp. 255-276). Montrouge : INSEE.
- INE – Instituto Nacional de Estatística (2019). Sustainable Development Goals - Indicators for Portugal. 2030 Agenda. Lisbonne : INE.
- Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale (2019a). The territorial dimension in SDG indicators: geospatial data analysis and its integration with statistical data. Lisbonne : INE.
- Comité régional Europe de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale (2019b). The integration of statistical and geospatial information — a call for political action in Europe. Luxembourg : Office des publications de l'Union européenne.
- CEE (2019). Étude approfondie de l'utilisation de l'imagerie satellitaire et des autres technologies d'observation de la Terre dans les statistiques officielles. Document établi par le Canada et le Mexique pour la soixante-septième réunion plénière, Genève, 26-28 juin.
- Nations Unies (2015). Transformer notre monde : le Programme de développement durable à l'horizon 2030. Résolution A/RES/70/1 adoptée par l'Assemblée Générale le 25 septembre 2015.