

**Conseil économique et social**

Distr. générale
13 juillet 2015
Français
Original : anglais

Commission économique pour l'Europe**Conférence des statisticiens européens****Groupe d'experts des recensements de la population
et des habitations****Dix-septième session**

Genève, 30 septembre-2 octobre 2015

Point 4 de l'ordre du jour provisoire

Statistiques par mailles : perspectives et difficultés

**Produits innovants pour la diffusion d'informations
géographiques (partie II)****Note de l'Institut national de la statistique de l'Espagne***Résumé*

Les informations géoréférencées utilisées dans le cadre du recensement de 2011 offrent de vastes perspectives. Au cours des derniers mois, nous avons continué à élaborer de nouveaux produits qui ont recours à ces informations et répondent aux demandes décelées auprès des utilisateurs. Le présent document, qui constitue la deuxième partie de celui qui a été présenté l'année dernière, reproduit en détail les faits les plus récents enregistrés.

De plus sont également expliqués certains aspects non mentionnés dans la première partie, comme les règles de confidentialité appliquées et l'algorithme qui subdivise le territoire en cellules dont chaque côté va jusqu'à 50 mètres selon la taille de l'échantillon de population.



I. Introduction

1. Les résultats obtenus dans tous les secteurs au cours de ces dernières années grâce aux informations géoréférencées sont incontestables. Il est de plus en plus courant de trouver des systèmes qui stockent les coordonnées des utilisateurs et présentent des informations sur des cartes pour une interprétation plus intuitive.

2. Les recensements, en tant qu'activité statistique qui fournit des renseignements très détaillés sur le territoire, ne pouvaient pas laisser passer l'occasion d'utiliser des informations géographiques. Voici quelques exemples qui montrent combien il est important de disposer d'informations géoréférencées lors d'un recensement :

- Une chaîne de supermarchés peut ainsi choisir l'emplacement d'un nouveau magasin;
- Une administration locale peut décider de l'emplacement le plus adéquat pour construire une école ou un aéroport;
- Il est possible de connaître des détails sur la population vivant dans un certain rayon.

II. Nouveaux produits géographiques diffusés

3. Dans le document présenté l'année dernière¹ figurait un ensemble de produits permettant de consulter des informations géographiques, tels que cartes, tableaux prédéfinis, fichiers de microdonnées, etc., relatifs au recensement. Tous ces produits pouvaient être consultés sur le microsite consacré au recensement de 2011².

4. Au cours de la présente année, nous avons continué à mettre au point de nouveaux produits et avons également pris en compte les avis des utilisateurs pour modifier les fonctions des produits existants.

5. Ainsi, les règles applicables à la confidentialité des informations ont été assouplies afin d'offrir aux utilisateurs un volume plus important de renseignements. On trouvera plus de précisions sur les nouvelles règles de confidentialité dans la section 4 de la présente note.

A. Cartes géographiques thématiques

6. L'année dernière, nous avons intégré dans notre système de diffusion des données recensées la possibilité de visualiser sept cartes géographiques à différents niveaux.

7. Les utilisateurs peuvent choisir directement sur la carte la zone géographique qu'ils veulent analyser. Selon le niveau de zoom, il est possible de voir trois cartes différentes :

- Cellules de 100 km de côté (10 000 km²);
- Cellules de 10 km de côté (100 km²);
- Cellules de 1 km de côté (1 km²).

¹ http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2014/mtg1/WP22_s5_Spain.pdf.

² http://www.ine.es/en/censos2011_datos/cen11_datos_inicio_en.htm.

Figure 2
Couverture de la publication



9. Cette publication contient 25 cartes qui ont été établies à l'aide d'informations concernant les variables de recensement les plus intéressantes :

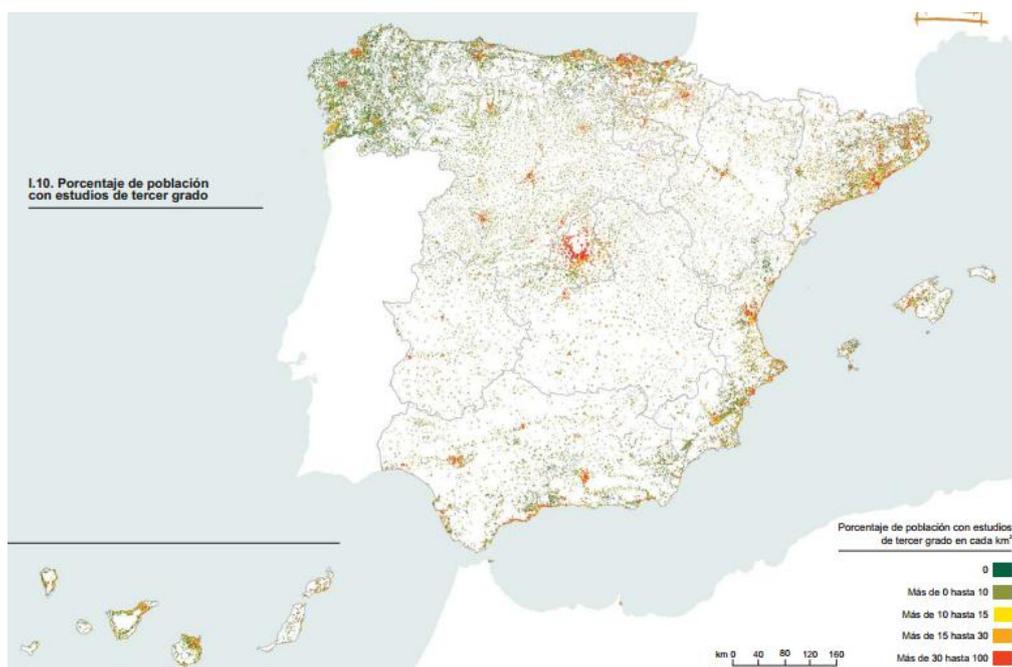
a) Population (14 cartes) : population totale, pourcentage de ressortissants étrangers, nationalité étrangère prédominante, âge moyen, pourcentage de personnes de 65 ans et plus, pourcentage de personnes âgées de 25 à 34 ans qui vivent avec leurs parents, pourcentage de personnes qui vivent à un endroit différent de leur lieu de naissance, année moyenne d'arrivée en Espagne, niveau d'études moyen, pourcentage de personnes ayant un niveau d'études élevé, nombre moyen d'enfants, temps moyen pour se rendre au lieu de travail ou d'étude, et pourcentage de personnes passant 14 nuits ou plus dans une autre municipalité;

b) Ménages (3 cartes) : taille moyenne, pourcentage de ménages d'une personne, pourcentage de ménages comptant deux générations et plus;

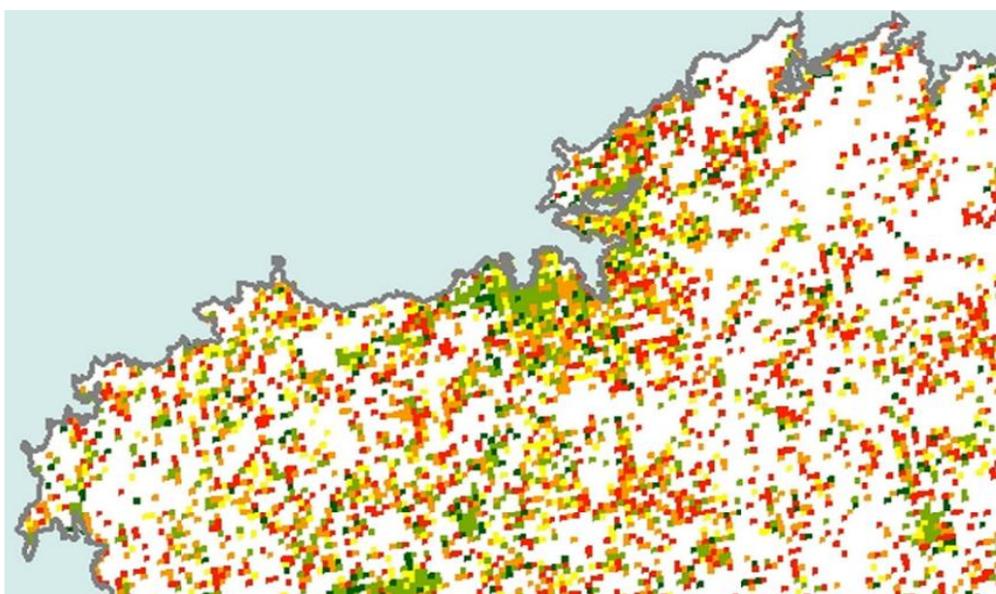
c) Logements (8 cartes) : pourcentage de logements secondaires ou vacants, surface utile moyenne, pourcentage de logements connectés à Internet, pourcentage de logements dotés d'un système de chauffage, pourcentage de logements loués, pourcentage de logements hypothéqués et durée moyenne de construction.

10. Ces cartes rendent compte des informations d'un point de vue différent et innovant (toutes les cellules étant de même taille, soit 1 km²) et indépendant du point de vue administratif auquel l'on est habitué. Dans la plupart des cas et, en particulier, dans les grandes villes, les cartes de 1 km² contiennent des renseignements plus détaillés et montrent mieux la réalité que les cartes municipales.

Figure 3
Exemple de cartes prédéfinies (pourcentage de personnes ayant fait des études supérieures)



11. L'un des avantages les plus importants de ces cartes est le fait qu'elles offrent la possibilité de visualiser des données détaillées en un coup d'œil. En outre, la publication est recommandée pour tous les types d'utilisateurs, pas forcément ceux qui connaissent bien les informations recensées. Il est très simple de comparer les données entre différentes parties de l'Espagne.



III. Cellules inférieures à 1 km² : algorithme de subdivision

12. La maille de 1 km² utilisée par l'Institut national espagnol de statistique (INE) pour la diffusion des informations géographiques a été établie par Eurostat³. Selon nos chiffres, 79 857 cellules comptent au moins un logement et des personnes vivent dans 63 527 de ces cellules.

13. Le site de l'INE consacré au recensement permet aux utilisateurs de consulter des informations concernant des cellules de 100 km², 10 km² et 1 km².

14. Cette maille peut être insuffisante, notamment en zone urbaine. Il convient de rappeler que les informations géoréférencées proviennent d'une enquête qui était un des éléments du recensement effectué en Espagne. L'enquête a porté sur près de 9 % de la population. Dans les cellules comptant un plus grand nombre d'unités d'échantillonnage (individus), nous avons procédé à des subdivisions supplémentaires (allant jusqu'à 50 m de côté) selon la taille de l'échantillon de population.

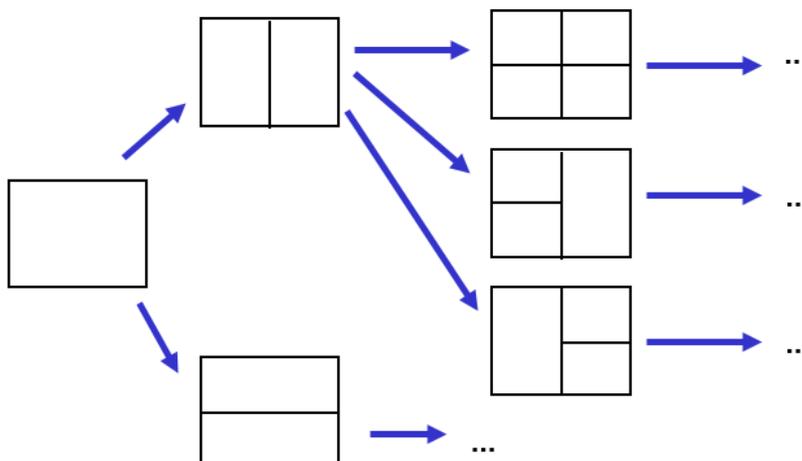
15. Les utilisateurs peuvent voir les subdivisions effectuées en suivant le lien ci-après (cliquer sur l'option de navigation illimitée) : <http://www.ine.es/censos2011/visor/>.

16. L'algorithme de subdivision des cellules de 1 km² vise trois situations distinctes :

- Situation 1 : le nombre d'unités d'échantillonnage de la cellule de 1 km² est inférieur à 40. Dans ce cas, aucune subdivision supplémentaire de la cellule n'est nécessaire.
- Situation 2 : le nombre d'unités d'échantillonnage de la cellule de 1 km² est égal ou supérieur à 40 et inférieur à 250. Dans ce cas, il sera appliqué de manière récursive des « demi-subdivisions » (verticales et horizontales) qui laissent au moins 20 unités d'échantillonnage des deux côtés.

Figure 5

Mécanisme de subdivision des cellules



³ On trouvera plus de précisions sur la maille établie par Eurostat en suivant le lien ci-après : http://www.efgs.info/data/eurogrid/Grid_ETRS89_LAEA_ES_1K.zip/at_download/file

- **Situation 3** : le nombre d'unités d'échantillonnage de la cellule de 1 km² est égal ou supérieur à 250. Dans ce cas, la cellule de 1 km² est tout d'abord subdivisée en 400 sous-cellules de 50 m de côté. On calcule ensuite le paramètre « p » :

$$p = 20 * \left[\left(\sqrt{\frac{\text{unités_d'échantillonnage}}{20}} \right) \right] + 1$$

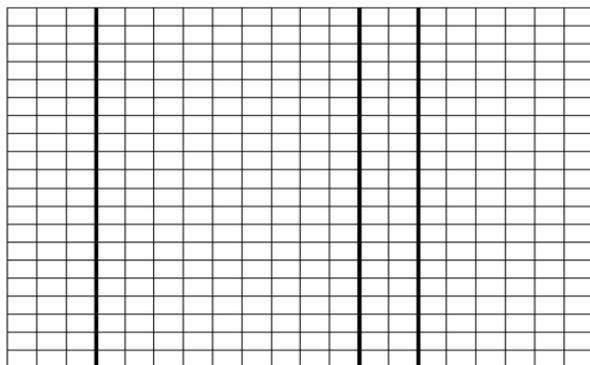
17. La cellule de 1 km² dont la densité est la plus élevée en Espagne⁴ comptait 3 209 unités d'échantillonnage (soit 52 845 habitants environ) et était subdivisée en 273 sous-cellules.

18. Enfin, les trois tâches suivantes ont été réalisées :

- Tracer autant de lignes verticales que possible qui laissent au moins « p » unités d'échantillonnage à gauche et à droite de chaque subdivision.

Figure 6

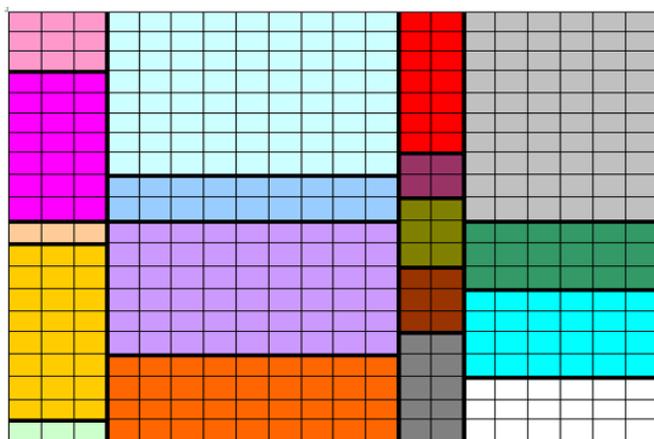
Première étape (n = 500, p = 120), quatre subdivisions



- Tracer autant de lignes horizontales que possible qui laissent au moins « 20 » unités d'échantillonnage en haut et en bas de chaque division.

Figure 7

Deuxième étape. 18 subdivisions



⁴ Située à L'Hospitalet del Llobregat (Barcelone).

c) Tracer autant de lignes verticales que possible qui laissent au moins « 20 » unités d'échantillonnage à gauche et à droite de chaque subdivision.

Figure 8
Troisième étape. 20 subdivisions

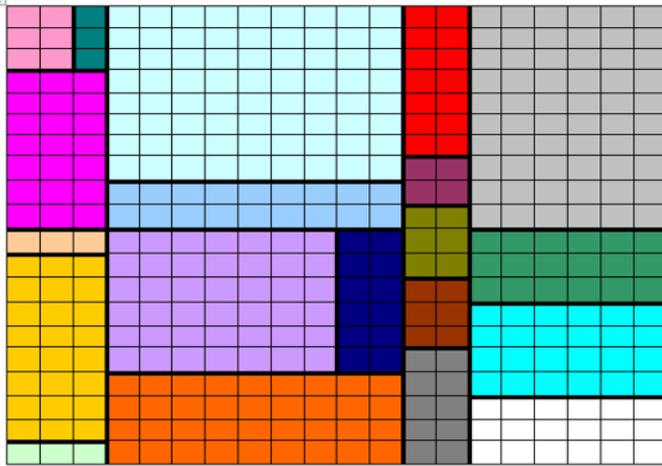


Figure 9
Exemple d'algorithme utilisé pour la ville de Barcelone



IV. Traitement de la confidentialité

19. L'un des aspects les plus importants de la diffusion d'informations géoréférencées est le traitement de la confidentialité des données.

20. L'INE a appliqué les mêmes mesures concernant la confidentialité des produits géographiques que pour d'autres types de produits. Les informations étaient diffusées lorsque l'une des conditions suivantes étaient remplies :

a) Lorsque le quotient du nombre d'unités d'échantillonnage existant et du nombre de catégories⁵ figurant dans la requête est supérieur ou égal à 5, la totalité des informations demandées est diffusée.

b) Autrement, lorsqu'il y a au moins cinq unités d'échantillonnage dans l'une quelconque des ventilations figurant dans la requête, les informations concernant ces ventilations qui satisfont à la condition susmentionnée sont diffusées.

21. L'exemple ci-après montre clairement comment cette règle fonctionne. Imaginons qu'un utilisateur souhaite obtenir des informations concernant la situation matrimoniale légale (5 catégories) dans une zone qui contient 20 unités d'échantillonnage.

a) Premièrement, le quotient de 20 et de 5 sera calculé. Le résultat, à savoir 4, n'est pas supérieur à 5, de sorte que la totalité des informations demandées n'est pas diffusée.

b) Deuxièmement, imaginons que ces 20 unités d'échantillonnage soient réparties comme suit : 7 célibataires, 8 personnes mariées, 2 veufs et 3 divorcés. Dans cette situation, les informations ayant trait aux personnes célibataires et mariées seront diffusées car le nombre des unités d'échantillonnage de ces ventilations est supérieur à 5.

22. En outre, pour essayer de préserver la confidentialité des informations diffusées et d'empêcher une divulgation de données au moyen des différences, toutes les données sont arrondies au multiple entier de 5 le plus proche (par exemple, si le résultat d'une requête est 422, le nombre diffusé sera 420).

23. Enfin, quant aux cartes thématiques comportant des informations géographiques, elles ont toujours été élaborées avec mention de plages d'intervalle dans les légendes de sorte que des informations individuelles ne sont jamais dévoilées.

24. Toutes les mesures effectuées pour protéger les informations garantissent que, malgré des niveaux géographiques très détaillés, 1 km² par exemple, des informations individuelles ne seront jamais révélées.

V. Conclusion

25. Les informations géoréférencées fournies par un recensement sont très variées. Au contraire du niveau administratif, très hétérogène au niveau des surfaces et où il est difficile de procéder à des comparaisons homogènes, il est possible, à l'aide des informations géoréférencées du système de recensement de l'INE, de comparer des unités qui sont toujours de même surface ou de définir des régions pour réaliser une analyse d'une manière presque illimitée.

26. De plus, les informations géoréférencées sont aussi un outil très intéressant pour analyser la situation dans les régions frontalières à l'intérieur d'un pays ou d'un pays à l'autre.

27. Certaines institutions ont pris acte de la puissance des informations géoréférencées⁶. Eurostat essaie d'encourager les pays à les collecter (bien qu'elles ne puissent pas être incorporées dans le règlement relatif au recensement de 2021).

⁵ Pour quelques variables sensibles, il existe un « facteur de sensibilité » supérieur à 1 et la condition à remplir est que le quotient du nombre d'unités d'échantillonnage existant et du produit du nombre de catégories par le facteur de sensibilité des variables concernées par la requête soit supérieur ou égal à 5.

28. Toutes ces raisons expliquent pourquoi il y aura très probablement un accroissement notable du nombre de pays qui projeteront de diffuser des informations géographiques au cours du prochain recensement.

⁶ Par exemple, il est prévu d'utiliser ces informations comme source de données pour répartir le budget des fonds européens. De plus, d'autres institutions telles que l'EFGS (<http://www.efgs.info/>) ont consacré des années à promouvoir l'élaboration de statistiques par mailles.