

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

Genève, 22-24 juin 2020

Point 5 a) 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****Géographes et Statisticiens nationaux en France :
deux décennies de synergies croissantes*****Note de l'Institut national de la statistique et des études économiques
(Insee) et l'Institut national de l'information géographique et forestière
(IGN)***Résumé*

Afin de faciliter la réponse aux besoins croissants d'information statistique finement localisée, de nombreuses initiatives ont été prises au cours des dernières années, que ce soit au niveau global, régional ou national. Elles se donnent comme objectif commun une meilleure intégration des informations statistiques et géographiques. Un outil privilégié pour y parvenir est, notamment, la promotion d'une collaboration accrue entre les instituts nationaux de statistique (INS) et ceux de géographie (NMA). La France, que ce soit par l'intermédiaire de son Institut National de la Statistique et des Études Économiques (Insee) et/ou de son Institut national de l'information géographique et forestière (IGN) participe pleinement aux réflexions internationales. Elle y apporte son expérience, mais en tire également profit dans le but rationaliser sa propre gestion des faits géographique et statistique. Dans ce document, nous présentons dans une première partie les contributions spécifiques de la France aux réflexions internationales. Dans une seconde partie nous montrons, dans une perspective historique, comment l'Insee s'est organisé et va s'organiser pour intégrer l'information géographique dans ses processus de production.

Le présent document est soumis pour examen au séminaire de la Conférence des statisticiens européens sur le thème « Nouveaux rôles des organismes nationaux de statistique et d'information géospatiale dans les récents écosystèmes nationaux de données ».

* Le présent document a été soumis en retard en raison de la transmission tardive de renseignements provenant d'autres sources.



I. Géographie et statistique au niveau international, les contributions françaises

A. Au niveau global – l’Initiative des Nations Unies sur la gestion de l’information géospatiale à l’échelle mondiale (UN-GGIM)

1. Au niveau global les questions économiques et sociales sont relatives au réchauffement climatique, aux crises migratoires ou sanitaires, au développement durable. Leur description ou la prise de décisions politiques qu’elles impliquent n’est possible, de manière efficiente, que si une information géographique de référence est disponible de manière cohérente et interopérable pour l’ensemble des pays. La mise en place d’une telle interopérabilité est l’un des objectifs de la Gestion Mondiale de l’Information Géospatiale (UN-GGIM) établie par le conseil économique et social de l’ONU en 2013. Le comité d’expert de l’UN-GGIM a cependant rapidement constaté que la valorisation de l’information géographique serait d’autant plus élevée, qu’elle est associée à d’autres types d’information, dont l’information statistique. C’est pourquoi il a mis en place, dès 2013, un groupe d’experts sur l’intégration des informations statistiques et géospatiales (EG-ISGI)¹, dont les membres sont recrutés parmi les instituts nationaux de statistiques et de géographie. La production majeure de ce groupe est un cadre statistique et géospatial global (GSGF²) qui au travers de 5 principes, fournit des recommandations pour une gestion efficiente d’un système d’information statistique adossé à un système d’information géographique. Ces recommandations servent de référence mondiale et irriguent ensuite les travaux à dimension régionale ou nationale ou relatifs à des domaines spécifiques, comme ceux du groupe de travail d’inter-agences sur l’information spatiale pour les Objectifs de Développement Durable³.

2. L’Insee est membre actif de l’EG-ISGI depuis sa création. Sa contribution spécifique est liée à son expérience reconnue dans la gestion de la confidentialité des données statistiques finement localisées⁴. A ce titre, dans le cadre de l’EG-ISGI, l’Insee pilote une *task team* sur la confidentialité et le respect de la vie privée lors de la diffusion de résultats statistiques à un niveau géographique fin. Les conclusions de ces travaux collectifs avec l’Allemagne, la Finlande, la Namibie, la Nouvelle-Zélande, le Mexique et les USA doivent être diffusés à la fin de l’année 2021.

3. La contribution spécifique de l’IGN s’inscrit quant à elle dans le cadre de l’UN-GGIM Europe, qui porte les mêmes enjeux que l’UN-GGIM mais déclinés au niveau continental. L’IGN pilote ainsi un groupe de travail visant à définir le socle minimal de données géographiques, ainsi que leur contenu, dont la mise à disposition pan-européenne permet de soutenir la mise en œuvre des politiques publiques de développement durable de manière cohérente entre pays. Le groupe de travail a porté son attention de manière plus spécifique sur les thèmes intéressant les instituts nationaux de statistique, tels que *adresses, bâtiments et parcelles cadastrales*. Ces thèmes ont en effet été identifiés comme étant de nature à faciliter le travail d’intégration des informations géographique et statistique.

B. Au niveau européen

4. Les réflexions de l’UN-GGIM et de l’UN-GGIM Europe servent de cadre de référence aux travaux d’Eurostat pour répondre à ses besoins propres en matière de géographie. Afin de favoriser la collaboration des différents instituts de statistiques et géographique européens, Eurostat a ainsi pris différentes initiatives :

¹ Expert Group on Integration of Statistical and Geospatial Information

² Global Statistical Geospatial framework

³ IA EG working group in geospatial information

⁴ Les travaux de l’Insee en la matière ont été récompensés par le premier prix du jeune statisticien lors de la conférence de l’International Association of Official Statistics. Ils font également l’objet d’un chapitre particulier dans le cadre du manuel d’analyse spatiale décrit ci-après.

- Eurostat a mis en place des réunions annuelles officielles regroupant les deux communautés et des représentants des organisations internationales, dont la CEE-ONU.
- Eurostat finance par ailleurs des projets multi-pays (appelés ESSnet) visant à faciliter l'intégration des informations géographiques et statistiques : les projets Geostat 1 à 4. Le premier de la série étudiait les questions méthodologiques liées à la diffusion de données carroyées. Geostat 4, qui est en cours, étudie notamment les conditions de mise en œuvre du GSGF au niveau européen. Il vise également à définir un cadre de référence qualité pour la production et la diffusion de statistiques finement localisées.
- Eurostat finance les initiatives nationales en termes d'intégration de la géographie par un système de subventions spécifiques.
- Eurostat finance les activités du Forum Européen de Géographie et de Statistique (EFGS). L'EFGS rassemble également, dans le cadre d'une association fonctionnant sur la base du volontariat, les membres des deux communautés. Sa conférence annuelle est depuis quelques années un lieu de riches échanges entre, désormais, plus de 200 participants.

5. L'Insee a participé à l'ensemble des projets Geostat (sauf le 3). Au sein du projet Geostat 4, qui est en cours, l'Insee contribue activement aux travaux de rapprochement entre le GSGF et les modèles de description des processus statistiques produits par l'Unece notamment : GSBPM, GAMSO, CSPA, GSIM.

6. L'Insee produit des statistiques et réalise des travaux d'études fondés sur les données qu'il produit. Afin de faire bénéficier l'ensemble de la communauté de cette spécificité de l'Insee par rapport à d'autres instituts nationaux de statistique (INS), Eurostat a confié à l'Insee et a co-financé la rédaction d'un manuel de statistiques spatiales. Ce travail a également reçu le soutien de l'EFGS. A la différence des autres initiatives, qui étudient les conditions de mise en place ou de diffusion d'un système d'information statistique finement géoréférencé, le manuel part du principe de l'existence d'un tel système. En 14 chapitres, le manuel décrit différentes méthodes de traitement de statistiques spatiales mobilisables destinées à enrichir ou faciliter les opérations de production, de diffusion ou d'analyse statistique. Spécifiquement destiné aux INS, le manuel met ainsi l'accent sur des questions, peu abordées dans les équivalents universitaires, comme l'échantillonnage spatial ou la confidentialité statistique et données géographiques. Ce manuel diffusé gratuitement en Français et en Anglais sur les sites de l'Insee, d'Eurostat et de l'EFGS a été téléchargé plus de 10 000 fois en deux ans.

7. Enfin, l'Insee et l'IGN ont organisé conjointement la conférence de l'EFGS à Paris en novembre 2016. En plus des thèmes de collaboration entre les deux communautés, l'agenda de la conférence a été particulièrement orienté vers l'analyse spatiale et l'ouverture au monde académique.

II. Géographie et Statistique au niveau national

8. Afin de répondre à ses besoins, l'Insee gère historiquement en interne son propre système d'information géographique ainsi que l'intégration de ce dernier dans le système d'information statistique. Au sens du projet Geostat 2 ce type d'organisation est qualifié de « *In house* ». La mise en place d'une collaboration récente mais néanmoins fructueuse avec l'IGN, n'a pas remis en cause les principes structurants de cette organisation. Les orientations actuelles de l'IGN, en cohérence avec les principes du GSGF permettent d'entrevoir à moyen terme un partage des rôles plus adaptés aux missions spécifiques de chaque institut.

A. L'organisation actuelle à l'Insee

1. Jusqu'en 1999

9. Le recensement de la population est la principale production statistique consommatrice d'information géographique. En France, en l'absence de registre national de

population localisée, le recensement est de type « *traditionnel* ». Les citoyens remplissent un questionnaire auto-administré délivré par des agents recenseurs.

10. L'organisation du travail des agents nécessite de leur assigner une zone de collecte ayant un nombre de logements adapté à la charge de travail et des limites géographiques facilement identifiables sur le terrain. Jusqu'en 1999, ces zones de collecte étaient dessinées à la main sur des cartes IGN papier dans les directions régionales de l'Insee. Dans les plus grandes villes ces zones de collecte étaient assemblées en zones un peu plus grandes (comprenant de l'ordre de 2000 habitants), appelées *IRIS*, pour les besoins de diffusion des résultats statistiques du recensement.

11. Une gestion méticuleuse de la codification des zones de collecte et de celles de diffusion conduisait à ne pas avoir recours à un système d'information géographique précis, qui n'existait pas par ailleurs.

2. De 1999 à 2016

12. Le rythme des recensements français était décennal jusqu'en 1999. Afin de réduire les délais de production et de diffusion du recensement, l'Insee est passé en 2004 à un système d'enquêtes annuelles de recensement. Dans les grandes communes, plus de 10 000 habitants, 8 % des adresses d'habitation sont recensées aléatoirement chaque année. Un cinquième des petites communes sont recensées exhaustivement chaque année. Le recensement annuel lui-même est fondé sur l'exploitation, glissante année après année, de 5 enquêtes annuelles consécutives, s'appuyant ainsi sur 40 % des adresses d'habitation des grandes communes et 100 % des petites communes. Le questionnaire est toujours auto-administré (papier ou web, au choix du répondant).

13. Dans les petites communes, le passage à des enquêtes annuelles de recensement ne change pas fondamentalement le besoin en information géographique. Des cartes permettent toujours de définir les zones de collecte. Dans les grandes communes, le changement est plus profond. Il implique l'existence d'un registre d'adresses d'habitation. En l'absence d'offre externe répondant aux besoins de l'Insee dans le calendrier de mise en place du projet de nouveau recensement, l'Insee a créé son propre répertoire d'adresses (le Répertoire d'immeubles localisés : RIL) ainsi que les conditions de sa maintenance annuelle dans le cadre d'un partenariat avec les mairies. Les adresses du registre sont décrites sémantiquement mais aussi géocodées afin d'aider à leur repérage sur le terrain ou les affecter automatiquement à des zones utiles pour la diffusion. En plus du répertoire d'adresses, l'Insee gère les contours géométriques des communes et des zones de diffusion pour les grandes communes.

14. L'existence du RIL sert dans un deuxième temps au géoréférencement de fichiers administratifs pour répondre à d'autres besoins de la statistique publique. En l'absence d'identifiant interopérable des adresses, ou de norme partagée sur la description des adresses, les adresses des unités statistiques du fichier à géoréférencer sont comparées sémantiquement à celles du RIL. La qualité du résultat dépend dès lors de la qualité du RIL en termes de fraîcheur et d'exhaustivité, de celle du moteur d'appariement, de celle de la saisie des adresses dans le fichier à géoréférencer et de la qualité générale de l'adressage en France. Sur ce dernier point, il existe en France de nombreuses habitations ayant des adresses non numérotées. C'est le cas dans les hameaux de zones rurales, par exemple. Afin d'assurer néanmoins la qualité statistique des résultats diffusés, une équipe de 15 personnes est chargée de traiter manuellement les cas où l'appariement est de mauvaise qualité quelle que soit l'origine du défaut d'appariement.

3. Depuis 2016

15. Le gouvernement français a confié à l'IGN en 1999 l'élaboration puis la production d'un Référentiel à Grande Échelle (RGE®) décrivant le territoire national et l'occupation de son sol de façon précise, complète et homogène. Disponible depuis 2006, le RGE® est composé de cinq composantes (orthophotographique, topographique, altimétrique, parcellaire et adresse) qui fournissent des informations géographiques parfaitement superposables. Depuis le 1er janvier 2011, le RGE® est diffusé au seul coût de reproduction

et de diffusion des données concernées, dès lors que celles-ci sont utilisées pour l'exercice d'une mission de service public ne revêtant pas un caractère industriel ou commercial.

16. La date de première de diffusion du RGE® 2006 et plus encore ses modalités de diffusion jusqu'en 2011 ont été des freins à son utilisation dans le cadre des activités de l'Insee. La première refonte d'ampleur du recensement tournant à partir de 2013 a été l'occasion de revoir le partenariat Insee – IGN.

17. Le premier constat dressé par l'Insee a été que sa gestion interne du fait géographique dans le RIL et le répertoire de contours associé conduisait à des incohérences topographiques quand les informations du RIL étaient représentées sur des fonds du RGE®. Par exemple, des adresses du RIL de Saint-Malo (en Bretagne) se retrouvaient dans la mer pour le RGE®. Afin de traiter ces incohérences, l'Insee et l'IGN ont mis en place un premier partenariat visant à caler les informations géographiques de l'Insee, que ce soit la position des adresses ou les contours des IRIS, sur les informations du RGE®. Dans un second temps, afin d'assurer dans la durée la cohérence retrouvée de leurs références géographiques, l'Insee et l'IGN ont signé deux conventions.

18. La première convention confère à l'IGN la gestion entière des contours géométriques nécessaires à l'Insee, dont celle des IRIS. La seconde met en place les conditions de l'échange mutuellement profitable suivant. L'IGN met à disposition de l'Insee les couches et les services cartographiques nécessaires à un positionnement, par l'Insee, des nouvelles adresses d'habitation repérées dans le cadre de son partenariat avec les mairies, qui soit cohérent avec le RGE®. L'IGN intègre également dans un délai de 10 jours les nouvelles voies repérées par l'Insee. En retour, l'Insee fournit à l'IGN son stock d'adresses d'habitation localisées et le flux d'adresses nouvelles.

19. On notera que cette organisation reste principalement interne. L'Insee continue de gérer son répertoire d'adresses en intégrant notamment des adresses nouvelles, qu'il repère et localise de façon cohérente cependant avec le RGE®. Cette externalisation partielle tient au fait que la composante adresse du RGE® ne répondait pas, en 2013, aux spécifications techniques de l'Insee, en termes de fraîcheur, d'exhaustivité, d'interopérabilité des identifiants d'adresses, de capacité à séparer adresses d'habitation de celles où s'exercent exclusivement des activités économiques par exemple.

20. En parallèle naissait le projet BAN (Base adresse nationale). La base adresse nationale (BAN) est le fruit d'un partenariat inédit entre l'IGN, la mission Etalab, le groupe La Poste, la Direction Générale des Finances Publiques (DGFIP) et l'association OpenStreetMap France. Cette base de données libre et collaborative a pour but de référencer la position géographique de l'intégralité des adresses réparties sur le territoire français. Finalement produite en 2018 par la direction interministérielle du numérique, l'IGN et La Poste, elle est alimentée en continu par les collectivités locales, les services de secours, les institutions telles que la DGFIP et l'Insee ou l'IGN, les entreprises et les citoyens. Un guichet collaboratif est mis à disposition des mairies par l'IGN pour permettre une intégration directe des données des mairies dans la BAN. En parallèle des imports massifs de données venant notamment de l'INSEE mais aussi de la Poste et de la Direction des finances publiques viennent contribuer à l'entretien de la BAN. Depuis le 1er janvier 2020, la BAN est diffusée sous licence ouverte. Une licence conçue par l'État afin de faciliter et d'encourager la réutilisation des données publiques mises à disposition gratuitement.

B. Les perspectives

1. Des données géographiques d'autorité et des procédures de géoréférencement communes

21. Le premier principe du GSGF vise à assurer que des résultats statistiques observés localement à partir de sources différentes révèlent bien des phénomènes sociaux ou économiques sous-jacents et non des artefacts liés à des différences dans les données géographiques mobilisées par source ou à des différences de méthodes de géoréférencement. Il promeut en conséquence l'utilisation de données géographiques d'autorité ainsi que des

méthodes d'intégration de la géographie qui soient communes à l'ensemble des processus de production statistique.

22. En cohérence avec ce principe, la France a entrepris en 2018 une réflexion sur les données géographiques souveraines. Considérant comme souveraines les données géographiques servant de support direct aux décisions de la puissance publique, un rapport parlementaire⁵ remis au gouvernement en 2018, a émis différentes recommandations sur ces données souveraines, leurs modalités de production, de diffusion, leurs possibles synergies, ainsi que sur les processus de gouvernance permettant une meilleure fédération des acteurs.

23. Le rapport parlementaire soutient l'enjeu d'une Base Adresse Nationale mais en reconnaît les limites et va jusqu'à préconiser d'étudier, en concertation avec les associations représentatives des communes, l'opportunité de systématiser l'obligation de réaliser des plans d'adressage et de rendre l'adresse opposable. Ces éléments sont de nature à harmoniser et rendre plus efficient les processus de production statistique. En conséquence, l'Insee, en partenariat étroit avec l'IGN, suit de manière rapprochée les évolutions de la BAN et y concourt activement.

24. En retour, ces perspectives ont conduit l'Insee à inscrire à son programme de travail une refonte d'ampleur de la gestion du fait géographique, ancrée sur la BAN. L'ensemble des productions statistiques pourraient être à terme concernées : recensement, géoréférencement des fichiers administratifs, répertoires des entreprises... En accord avec les principes du GSGF, l'objectif est que l'intégration des données géographiques soient une composante clairement identifiée des processus de production statistique au sens du GSBPM et qu'elle s'appuie sur une infrastructure informatique adaptée aux enjeux de rationalisation et d'efficacité, notamment par le développement d'API qui permettent l'interrogation directe des références géographiques ou statistiques, plutôt que via des copies locales de ces références.

2. D'autres perspectives de collaboration

25. Au-delà des questions d'infrastructure, d'autres pistes de collaboration ont émergé récemment et pourraient être systématisées.

26. L'IGN assure la maîtrise d'œuvre du portail national du territoire français (<http://geoportail.gouv.fr>). A ce titre, il a mis en ligne la représentation cartographique associée à la production par l'Insee de données carroyées sur la population et les revenus (voir figure 1). Cette production est pour l'instant limitée à une seule source statistique. Elle devrait être étendue à l'avenir aux données du recensement 2021, en cohérence avec le règlement européen sur le Censur. Les données sur l'emploi, les entreprises sont également candidates à une diffusion sous la forme de données carroyées.

27. Cette diffusion de données statistiques sous forme cartographique sera d'autant plus aisée que les deux instituts diffusent leur production selon le modèle Resource Description Framework (RDF). Ce modèle permet de décrire formellement les sources et leurs métadonnées, afin d'en permettre un traitement automatique et une certaine interopérabilité.

28. L'Insee finalise actuellement la mise au format RDF des nomenclatures territoriales dont elle a la responsabilité. L'Insee et l'IGN sont par ailleurs en contact pour la mise au même format des contreparties géométriques des nomenclatures territoriales. La finalisation de ces travaux permettrait de rendre plus efficace et dynamique la recherche de données statistiques et géographiques, en accord avec les concepts du *Linked Open Data*.

III. Conclusion

29. Du fait de l'évolution de l'offre de données géographiques, et après avoir été entièrement interne, la gestion du fait géographique à l'Insee s'est récemment orientée vers un rapprochement avec l'IGN, sans remettre pour autant en cause les fondements de l'organisation. Les projets structurants de mise en place de répertoire d'adresses, dans un premier temps, conduisent l'Insee à repenser plus en profondeur sa gestion de la géographie

⁵ V. Faure-Muntian (2018) : Les données géographiques souveraines, rapport au Gouvernement.

et à l'inscrire dans une dynamique de collaboration plus étroite. Sur cette dynamique de création d'infrastructure s'inscrivent, pour le bénéfice mutuel des deux communautés, d'autres opportunités de collaboration et notamment de diffusion de données statistiques finement localisées.

Figure
Mise en ligne par l'IGN de données carroyés Insee sur le portail national du territoire français (part des ménages pauvres)

