

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

Paris, 26-28 juin 2019

Point 2 b) de l'ordre du jour provisoire

Nouvelles sources de données – accessibilité et utilisation**Deuxième séance : aptitudes requises pour exploiter
les nouvelles sources de données****Amélioration de l'intégration des données à l'aide
du cadre statistico-géospatial mondial****Note de Statistics Sweden***Résumé*

Les données géospatiales et leur intégration dans la production statistique gagnent en importance car les organismes nationaux de statistique répondent aux besoins des utilisateurs de statistiques quant à l'accessibilité, à la dimension urbaine ou rurale, ou aux statistiques au niveau local en général.

Statistics Sweden a procédé à une évaluation nationale du cadre statistico-géospatial mondial, qui a été élaboré dans le cadre de l'Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale et approuvé par la Commission de statistique de l'ONU. L'évaluation porte sur les cinq principes du cadre du point de vue des capacités. S'agissant de l'intégration des données statistiques et géospatiales, les capacités s'appuient sur des personnes, des systèmes, des processus, des informations, des méthodes, des normes et des cadres ainsi que sur des mécanismes institutionnels.

Le présent document étudiera principalement la manière dont les pays peuvent améliorer leurs capacités d'intégration des données en utilisant un cadre mondial associé à un modèle mondial, l'Architecture commune de la production statistique. L'évaluation a permis à Statistics Sweden de déterminer les domaines où des améliorations sont possibles s'agissant des différents éléments des capacités. Elle a également contribué à mettre au jour les domaines où la situation est assez satisfaisante.

Ce document est présenté, pour examen, au séminaire de la Conférence des statisticiens européens de 2019 intitulé « Nouvelles sources de données – accessibilité et utilisation », la deuxième séance étant consacrée aux « aptitudes requises pour exploiter les nouvelles sources de données ».



I. Introduction

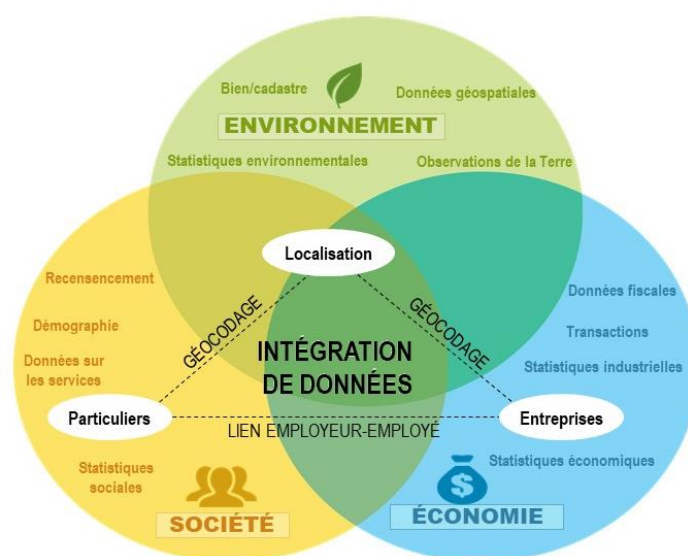
1. Les politiques territoriales appliquées aux niveaux national et européen, ainsi que le Programme de développement durable à l'horizon 2030¹, ont pesé sur les organismes de statistique nationaux européens, les obligeant à produire des statistiques pertinentes en temps voulu². Le Programme 2030 vise à ne laisser personne de côté ; d'un point de vue géographique, cela signifie par exemple qu'il faut prendre en compte les dimensions urbaine et rurale, ainsi que les questions d'accessibilité.
2. Il est plus important que jamais d'établir une corrélation entre les statistiques officielles et les réalités auxquelles les personnes sont confrontées, par exemple, dans les quartiers défavorisés. De plus, des statistiques au niveau local sont indispensables à la gestion des risques de catastrophe ou à la mise en place de mesures visant à améliorer l'accessibilité aux transports publics, aux écoles, aux centres de soins ou aux espaces verts. Le système statistique européen³ a reconnu cette nécessité en finançant les projets GEOSTAT⁴, GEOSTAT 2 par exemple, donnant ainsi aux organismes nationaux de statistique des orientations sur les éléments fondamentaux d'une infrastructure statistique plus souple fondée sur des données ponctuelles⁵.
3. Tous les pays de l'Union européenne produisent périodiquement des statistiques régionales, généralement à l'aide de la classification des unités administratives de la nomenclature des unités territoriales statistiques (NUTS) établie dans tous les États membres⁶. La base de données d'Eurostat fournit des statistiques régionales harmonisées. Cela étant, la situation est loin d'être harmonisée pour les utilisateurs à la recherche de plus de détails géographiques. Pour améliorer cette situation, on a introduit dans un premier temps l'utilisation de mailles statistiques, du degré d'urbanisation, etc., ainsi que la classification des unités administratives locales en tant que nouveaux produits géographiques, conformément à la version révisée du règlement de l'UE relatif aux typologies territoriales (règlement Tercet)⁷.
4. Les mailles statistiques sont appelées à devenir l'une des zones de diffusion au cours du prochain cycle européen de recensement de la population et du logement, qui aura lieu en 2021⁸. La production de statistiques fondées sur des mailles passe impérativement par la modernisation de la manière dont les pays procèdent à leurs recensements et ouvre la voie à un système de statistiques fondé sur les registres.

II. Corrélation entre la population et les entreprises

A. Avantages de l'intégration statistico-géospatiale

5. Les avantages de l'intégration statistico-géospatiale commencent à être largement reconnus et le Comité des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale a créé un groupe d'experts de l'intégration des données statistiques et géospatiales (EG-ISGI)⁹ [9]. Statistics Sweden prend activement part aux travaux du Groupe d'experts et a également coordonné le projet ESSnet/GEOSTAT 3, qui portait sur une mise en œuvre européenne du cadre statistico-géospatial mondial¹⁰. Un projet ESSnet est un réseau de plusieurs organisations membres du système statistique européen (SSE) dont le but est de produire des résultats qui seront bénéfiques pour l'ensemble du SSE.
6. Le cadre statistico-géospatial mondial a été adopté en 2016 et contribue directement à l'intégration des données statistiques et géospatiales¹¹. Statistics Sweden a commencé à mettre en œuvre au niveau national le cadre statistico-géospatial mondial, qui s'est révélé utile pour la planification et le recensement des améliorations apportées à cet égard aux statistiques, ainsi qu'il est indiqué plus loin dans le présent document.
7. Le cadre statistico-géospatial mondial contribue à ajouter une localisation géographique aux données statistiques qui décrivent des particuliers et des entreprises. Il contribue aussi à organiser la gestion des données géospatiales et les procédures nécessaires pour obtenir des statistiques de qualité.

Figure 1
**Corrélation entre la société, l'économie et l'environnement
 grâce à l'intégration des données¹²**



8. En outre, le cadre statistico-géospatial mondial constitue une passerelle entre l'organisme officiel de statistique et les professionnels des données géospatiales. Les unités statistiques sont géocodées au moyen d'identificateurs géospatiaux tels que les parcelles cadastrales, les adresses ou les bâtiments. Les données sont ensuite agrégées et diffusées pour définir les limites géographiques dans la couche des limites (ou la couche géographique). Cela facilite l'intégration des données entre les différentes sources et constitue une passerelle entre les ensembles de données statistiques et les ensembles plus traditionnels de données géospatiales, grâce à l'utilisation de données géographiques communes et à une plus grande souplesse dans les produits géographiques.

B. Les cinq principes du cadre statistico-géospatial mondial

9. Le cadre statistico-géospatial mondial est un cadre de haut niveau, ce qui signifie qu'il n'est pas censé offrir un plan ou un modèle détaillé de mise en œuvre. Il donne en revanche des orientations sur ce qui devrait être disponible dans les pays, offrant une grande souplesse quant à la manière de procéder.

10. Pour aider les pays à mettre en œuvre le cadre, des documents d'appui ont été élaborés en 2018, à la fois au niveau mondial (par le Groupe d'experts de l'intégration des données statistiques et géospatiales) et au niveau européen (dans le cadre du projet GEOSTAT 3). Le guide de mise en œuvre¹³ du projet GEOSTAT 3 montre bien comment comprendre les principes d'un point de vue européen.

1. Principe 1 : Utilisation de l'infrastructure géospatiale fondamentale et du géocodage

11. L'objectif de ce principe est d'obtenir une référence ou un attribut de localisation normalisé de qualité. Le cadre recommande d'utiliser les informations géospatiales existantes fournies par les autorités géospatiales spécialisées (par exemple, adresse physique, parcelle cadastrale, bâtiment ou autre description de localisation) pour attribuer les coordonnées exactes et/ou une petite zone géographique ou une maille de référence normalisée à chaque unité statistique (c'est-à-dire au niveau des microdonnées ou des données unitaires).

2. Principe 2 : Données unitaires géocodées dans un environnement de gestion de données

12. Le cadre statistico-géospatial mondial recommande que la corrélation entre un géocode et chaque unité statistique dans un ensemble de données (c'est-à-dire une personne, un ménage, une entreprise, un bâtiment ou une parcelle/unité de terrain) se fasse dans un environnement de gestion de données. Le stockage continu et contrôlé, au niveau des versions, des géocodes de grande précision permet d'utiliser tout contexte géographique lors de l'établissement des données à diffuser à l'avenir (c'est-à-dire en agrégeant des données dans diverses grandes unités géographiques ou pour s'adapter à l'évolution des unités géographiques dans le temps).

3. Principe 3 : Utilisation d'unités géographiques communes pour la diffusion de statistiques

13. Pour comparer les ensembles de données provenant de sources différentes, le cadre statistico-géospatial mondial recommande de recourir à un ensemble commun d'unités géographiques pour l'affichage, la notification et l'analyse d'informations sociales, économiques et environnementales.

4. Principe 4 : Interopérabilité statistique et géospatiale – données, normes et processus

14. Les statisticiens tout comme les spécialistes des données géospatiales exploitent leurs propres modèles de données et leurs propres capacités en matière de métadonnées, dont l'application n'est cependant pas souvent universelle. Les statisticiens ont recours au modèle générique d'informations statistiques (GSIM), à l'échange de données et de métadonnées statistiques (SDMX) et aux mécanismes de la Data Documentation Initiative (initiative de documentation des données), tandis que les spécialistes des données géospatiales appliquent d'autres normes¹⁴.

15. Dans le milieu de la statistique, il est nécessaire d'intégrer des processus et des normes géospatiaux dans les processus statistiques de façon plus uniforme. En 2018, la CEE a révisé le modèle générique du processus de production statistique¹⁵, en améliorant notamment la description de l'utilisation des données géospatiales et des méthodes y relatives dans le processus de production statistique.

5. Principe 5 : Statistiques géospatiales accessibles et utilisables

16. Ce principe du cadre statistico-géospatial mondial met en évidence la nécessité de recenser ou, le cas échéant, d'élaborer des politiques, des normes et des lignes directrices qui favorisent la diffusion, l'accessibilité, l'analyse et la visualisation des informations géospatiales. Un aspect important de ce principe est de veiller à ce que les données soient accessibles à l'aide de mécanismes sécurisés qui non seulement protègent la vie privée et la confidentialité, mais permettent aussi d'accéder aux données pour entreprendre diverses analyses favorables à la prise de décisions.

III. Capacités d'intégration des données statistiques et géospatiales

A. Évaluation de la capacité de Statistics Sweden

17. Un organisme statistique doit avoir la capacité de mener une ou plusieurs activités. Selon l'Architecture commune de la production statistique, les éléments généraux requis sont les suivants : personnes, processus, méthodes, technologies, normes et cadres¹⁶. Dans le cadre du projet GEOSTAT 3, qui étudie la manière d'appliquer le cadre statistico-géospatial mondial en Europe, Statistics Sweden a procédé à une évaluation nationale de chacun des principes du cadre.

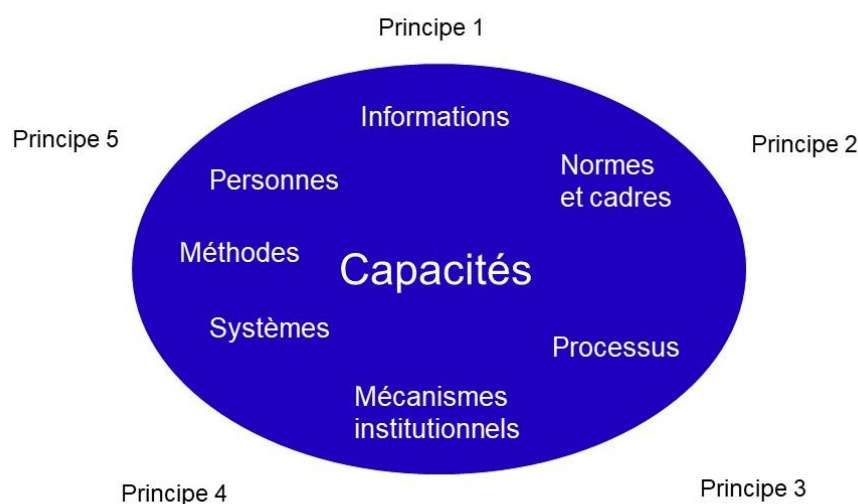
Tableau 1
Éléments des capacités

<i>Éléments des capacités</i>	
Personnes	Compétences, connaissances et aptitudes requises pour mener l'activité.
Systèmes	Applications TIC, matériel et plateformes requis par l'organisme pour mener l'activité.
Processus	Ensemble de mesures ou de tâches à privilégier par l'entreprise, pour ce qui est des méthodes, des normes et des systèmes selon que de besoin, pour mener l'activité de façon concrète et efficace.
Méthodes	Ensemble des techniques ou des algorithmes nécessaires pour mener l'activité.
Normes et cadres	Normes, cadres, guides et politiques nécessaires pour mener l'activité.
Informations	Données indispensables pour mener des activités d'intégration de données statistiques et géospatiales.
Mécanismes institutionnels	Infrastructures juridiques et institutionnelles requises pour mener et appuyer l'activité.

18. Il s'est révélé utile d'examiner le cadre du point de vue des capacités, étant donné que les principes du cadre doivent s'appuyer sur un certain nombre d'aptitudes, de processus, de méthodes, etc. L'évaluation a montré que, dans le cas de l'intégration statistique et géospatiale, les éléments clés des capacités peuvent être élargis pour inclure aussi des systèmes d'information et des mécanismes institutionnels¹⁷.

Figure 2

Éléments nécessaires à un institut national de statistique pour avoir des capacités d'intégration des données statistiques et géospatiales



19. L'évaluation peut aisément être menée par les instituts nationaux de statistiques qui ont récemment commencé à travailler à l'intégration statistique et géospatiale et par ceux qui cherchent à l'améliorer. Les mesures proposées dans le rapport suédois sont simples et prennent comme point de départ les besoins des utilisateurs.

20. L'examen des éléments clefs des capacités pour chaque principe a mis au jour les points forts et les points faibles de Statistics Sweden. Il est plus facile, en définissant les objectifs et en décidant des activités à privilégier, d'inclure des mesures dans les plans annuels et d'en assurer le suivi. Les mesures à prendre peuvent être classées dans la catégorie des nouvelles activités à entreprendre ou dans celle des activités à améliorer. L'évaluation a également révélé les activités à maintenir ou à réduire.

B. Nécessité de lancer de nouvelles activités

21. Statistics Sweden a recensé neuf mesures qui permettront d'améliorer les capacités d'intégration statistique et géospatiale. Il faut, par exemple, donner des orientations sur les pratiques en matière de données géospatiales et de géocodage similaires à ce que l'on trouve dans la documentation en ligne de l'Australian Bureau of Statistics¹⁸. Cela aidera à rationaliser les processus internes de Statistics Sweden et sera aussi utile aux autres autorités chargées d'établir des statistiques officielles. Nous avons également besoin d'une approche systématique concernant l'intégration géospatiale de tous les registres, le résultat souhaité étant une amélioration de la précision géospatiale. Aucune amélioration supplémentaire de la qualité n'étant nécessaire, cela permettra de gagner ultérieurement du temps au niveau des processus.

22. Il est important de fournir toutes les données statistiques géographiques sous forme de données ouvertes. Une autre mesure consiste à intégrer les données et les métadonnées géospatiales dans l'architecture de production statistique de Statistics Sweden. Cela permettra une meilleure intégration entre le processus de production statistique, la gestion des données, les métadonnées et les infrastructures.

23. Il nous faut également établir des objectifs concrets pour obtenir la qualité géographique souhaitable au niveau des registres (par exemple déterminer la part des lieux de travail dans le registre des entreprises pour assurer un bon géocodage). Nous devons aussi lancer des services cartographiques interactifs sur notre site scb.se et diffuser, sous la forme de données ouvertes, un certain nombre d'indicateurs de base concernant les petites zones géographiques statistiques.

C. Nécessité d'apporter des améliorations

24. Quatorze mesures ont été classées dans la « catégorie des améliorations à apporter » et visent les domaines dans lesquels nous travaillons déjà, mais dans lesquels nous pouvons faire plus pour améliorer nos capacités en matière d'intégration statistique et géospatiale.

25. Nous devons améliorer la communication de données à l'agence géospatiale nationale (Lantmäteriet) concernant le problème de la qualité dans la fourniture de données. Il convient également de mieux utiliser les déclarations sur la terminologie et la qualité des données convenues. Nous gagnerions à accroître l'utilisation du serveur SQL spatial pour mener des opérations spatiales. Il est nécessaire d'améliorer les informations relatives aux adresses figurant dans le registre des entreprises de façon à correctement géoréférencer tous les lieux de travail.

26. L'utilisation de données géospatiales dans l'ensemble du processus de production doit être améliorée en parallèle à une intensification de l'aide apportée par les spécialistes du système d'information géographique (SIG) au reste de l'organisme. Un exemple en est l'organisation de cours et d'ateliers de formation à l'utilisation des données géospatiales dans l'analyse statistique.

27. De plus, dans le cadre d'un service en ligne, nous devrions rendre toutes les données géographiques administratives et statistiques utilisées par les organismes qui relèvent du système national de statistique plus disponibles sous la forme de données ouvertes. Dans la mesure du possible, il faudrait aussi fournir des données rétrospectives pour faire en sorte que les statistiques rétrospectives puissent être correctement mises en correspondance avec des zones géographiques précises.

28. Nous devons également améliorer nos méthodes de protection de la confidentialité pour permettre aux utilisateurs de télécharger et de combiner divers ensembles de données avec plus de souplesse (différenciation géographique). Nous devrions aussi enregistrer une information cartographique dans la base de données statistique (pour compléter l'information existante dans les tableaux).

D. Maintien de ce qui fonctionne bien

29. L'évaluation montre également ce qui fonctionne déjà bien. La coopération est efficace entre Statistics Sweden et l'agence géospatiale nationale Lantmäteriet ainsi que d'autres fournisseurs de données géospatiales. Le système de statistique fondé sur des registres et des sources de données administratives constituent les bases de l'intégration.

30. Le service en ligne REGINA¹⁹ constitue une amélioration très utile de l'accessibilité à l'évolution historique des unités géographiques administratives. Il a été amélioré grâce à l'ajout de services de visualisation, mais devra être géré et perfectionné de manière à inclure un plus grand nombre d'unités géographiques administratives historiques.

31. Les connaissances en matière de services de recherche, de visualisation et de téléchargement exigées par l'Infrastructure d'information géographique dans la Communauté européenne sont précieuses lorsque nous étudions les moyens de renforcer nos services en ligne.

32. Nous disposons de processus automatiques de protection de la confidentialité qui fonctionnent bien, ce qui contribue à l'efficacité de la production de statistiques aréolaires en réduisant les risques de divulgation. Statistics Sweden dispose d'un portail qui fournit des statistiques à la demande et constitue un système de diffusion sécurisé pour les services commandés.

E. Réduction de l'inefficacité

33. L'évaluation des éléments des capacités a montré qu'il y aurait beaucoup à gagner à réduire les processus inefficaces concernant les contrôles de qualité et le chevauchement des activités. Il convient d'éviter la correction ponctuelle des données et les contrôles de qualité intervenant à la fin du processus de production, qui sont contraignants.

34. Il faut aussi mentionner le stockage inefficace des données et la duplication des bases de données. Statistics Sweden ne stocke plus les anciennes versions des données géospatiales. Celles-ci sont stockées dans un entrepôt de données national, ce qui rend la production plus complexe. Il existe également une confusion inutile quant aux différentes versions des zones géographiques (échelle, modification des limites, etc.), qui doit être réduite.

35. Le recours à davantage de données ouvertes permettra de réduire la nécessité de recourir à des services payants et les statistiques aréolaires seront plus accessibles et plus utilisées dans la société.

III. Conclusions

36. Les informations géospatiales sont encore considérées comme quelque chose qui concerne essentiellement un petit groupe d'experts du système d'information géographique de Statistics Sweden. L'évaluation vise les domaines dans lesquels de nouvelles initiatives et améliorations sont indispensables pour intensifier l'utilisation des données géospatiales et des processus connexes dans l'ensemble de l'organisme.

37. Pour chaque principe, un certain nombre d'éléments des capacités ont été évalués, ce qui a abouti à l'élaboration de propositions concernant ce que nous devons initier, améliorer, maintenir ou supprimer. Les activités recensées dans le cadre de l'évaluation ont depuis lors été incorporées dans les plans de travail annuels et des mesures importantes visant l'accès à davantage de données ouvertes ont été prises. Cela étant, il faut mener des

travaux supplémentaires pour atteindre l'objectif consistant à inclure les données géospatiales et les processus connexes et à faire en sorte qu'ils soient utilisés par l'ensemble de l'organisme.

38. Il est fortement recommandé que tout organisme national de statistique désireux d'améliorer ses capacités géospatiales procède à une évaluation. En mettant en œuvre le cadre statistico-géospatial et en mettant l'accent sur les éléments pertinents, l'organisme national de statistique peut hiérarchiser les domaines dans lesquels il faut investir en matière d'aptitudes, de processus et d'outils.

IV. Bibliographie

- ¹ Assemblée générale des Nations Unies. Résolution 70/1. Transformer notre monde : le Programme de développement durable à l'horizon 2030.
- ² Haldorson M., Zaccheddu P., Fohgrub B., Petri E. (2016). Geospatial information management in Europe – responding to the user needs, *Statistical Journal of the IAOS* 32.
- ³ Système statistique européen, <http://ec.europa.eu/eurostat/web/european-statistical-system/overview?locale=fr>.
- ⁴ Descriptifs des projets GEOSTAT, <http://www.efgs.info/geostat/>.
- ⁵ Moström J. (éditeur) et consorts (2017). A Point-based Foundation for Statistics, rapport final du projet GEOSTAT 2.
- ⁶ Règlement sur la nomenclature NUTS (2003). Règlement (CE) n° 1059/2003 du Parlement européen et du Conseil du 26 mai 2003 relatif à l'établissement d'une nomenclature commune des unités territoriales statistiques (NUTS), JO L 154, 21.6.2003.
- ⁷ Règlement Tercet, Typologies territoriales de l'Union européenne. <http://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies>.
- ⁸ Recensement européen, <https://ec.europa.eu/eurostat/web/population-and-housing-census/overview>.
- ⁹ Initiative des Nations Unies sur la gestion de l'information géospatiale à l'échelle mondiale – Groupe d'experts de l'intégration des données statistiques et géospatiales, <http://ggim.un.org/UNGGIM-expert-group/>.
- ¹⁰ Projet GEOSTAT/3, <http://www.efgs.info/geostat/geostat-3/>.
- ¹¹ Document de référence du Groupe d'experts de l'intégration des données statistiques et géospatiales de l'ONU (2016). Groupe d'experts de l'intégration des données statistiques et géospatiales : Global Statistical Geospatial Framework : Linking Statistics and Place, Current status and plans for development (juillet 2018).
- ¹² Illustration élaborée par Statistics Australia et Statistics New Zealand.
- ¹³ Cadre statistico-géospatial mondial Europe – Implementation guide for the Global Statistical Geospatial Framework in Europe, Proposal from the GEOSTAT 3 project, Version 1.0 (28 février 2019).
- ¹⁴ Documentation de l'atelier de la CEE consacré aux normes statistiques et géospatiales, Stockholm, novembre 2017, <https://www.unece.org/index.php?id=45404>.
- ¹⁵ Modèle générique du processus de production statistique, <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>.
- ¹⁶ Architecture commune de la production statistique, <https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture>.
- ¹⁷ Haldorson M., Moström J. (2018) : Implementing the Statistical Geospatial Framework at Statistics Sweden, rapport national établi dans le cadre du projet GEOSTAT 3.
- ¹⁸ Site Web de l'Australian Bureau of Statistics relatif à la géographie statistique, <http://www.abs.gov.au/geography>.
- ¹⁹ Service en ligne REGINA de Statistics Sweden. Informations disponibles en anglais dans un document PDF, <http://regina.scb.se/>.