United Nations ECE/CES/2020/28

Economic and Social Council

Distr.: General
8 May 2020
English
Original: French

Economic Commission for Europe

Conference of European Statisticians

Sixty-eighth plenary session
Geneva, 22–24 June 2020
Item 5 (a) of the provisional agenda
New roles for national statistical agencies and geospatial agencies in emerging national data ecosystems: Session 2: Experiences and results of concrete steps already taken by NSOs and the geospatial communities to modernize their role

Geographers and national statisticians in France: two decades of increasing synergies*

Note by the National Institute of Statistics and Economic Studies (INSEE) and the National Institute of Geographic and Forest Information (IGN)

Summary

In recent years, many initiatives have been taken to meet the growing need for highly localized statistics, at the global, regional and national levels. The common objective of these initiatives is to improve the integration of statistical and geographical information. One of the primary tools for achieving this is the promotion of increased collaboration between national statistical institutes and national mapping agencies. France is fully involved in international discussions, through both its National Institute of Statistics and Economic Studies (INSEE) and its National Institute of Geographic and Forest Information (IGN). The country shares its experience, while also benefiting from the discussions by streamlining its own management of geographical and statistical data. In the first part of this document, we set out the specific contributions of France to the international discussions. In the second part, we provide a historical perspective on the organization, past and future, by INSEE of the integration of geographical information into its business processes.

This document is presented to the Conference of European Statisticians seminar on "New roles for national statistical agencies and geospatial agencies in emerging national data ecosystems".

* This document was submitted late due to delayed inputs from other sources.

GE.20-06479 (E) 020620 020620

* 2 0 0 6 4 7 9 *

I. French contributions to geography and statistics at the international level

A. Global level – the United Nations Initiative on Global Geospatial Information Management (UN-GGIM)

1. At the global level, the economic and social issues at stake concern climate change, migration and health crises and sustainable development. It is only possible to describe them and make the necessary policy decisions efficiently if geographic reference information is consistently available and interoperable for all countries. Such interoperability is one of the goals of United Nations Global Geospatial Information Management, established by the Economic and Social Council in 2013. However, the Committee of Experts on Global Geospatial Information Management soon realized that geographical information becomes more valuable when it is combined with other types of information, including statistics. For this reason, in 2013 it established the Expert Group on the Integration of Statistical and Geospatial Information,[1] whose members are recruited from national statistical and geographical institutes. The main output of the Expert Group is the Global Statistical Geospatial Framework,[2] which provides recommendations for the efficient management of a statistical information system together with a geographical information system, based on five principles. These recommendations constitute a global standard and feed into work with a regional or national scope or focused on specific areas, such as the activities of the working group on geospatial information of the Inter-Agency and Expert Group on the Sustainable Development Goal Indicators.[3]

2. INSEE has been an active member of the Expert Group on Integration of Statistical and Geospatial Information since its establishment. Its specific contribution is related to its recognized experience in managing the confidentiality of highly localized statistical data.[4] On this basis, INSEE leads a task team on privacy and confidentiality in the dissemination of statistical results at a fine-grained geographical level. The conclusions of this joint work with Germany, Finland, Namibia, New Zealand, Mexico and the United States of America should be published at the end of 2021.

3. The specific contribution of IGN is made within the United Nations Committee of Experts on Global Geospatial Information Management for Europe (UN-GGIM: Europe), which has the same aims as UN-GGIM but at the continental level. To this end, IGN leads a working group that aims to define a minimum set of geographical data and data content that, when made available at the pan-European level, would support the consistent implementation of sustainable development policies across countries. The working group has focused more specifically on themes of interest to national statistical institutes such as addresses, buildings and cadastral parcels. These themes were identified as likely to facilitate the integration of geographical and statistical information.

B. European level

4. The discussions of UN-GGIM and UN-GGIM: Europe serve as a reference framework for Eurostat to meet its own geographical needs. Eurostat has taken various initiatives to promote cooperation between different European statistical and geographic institutions:

- Eurostat has established official annual meetings bringing together the two communities and representatives of international organizations, including the Economic Commission for Europe.

[1] EG-ISGI

[2] GSGF

[3] IAEG-SDGS

[4] The work of INSEE on this subject won the Prize for Young Statisticians of the International Association of Official Statistics. It is also covered in a dedicated chapter of the handbook of spatial analysis described below.

- Eurostat also funds multi-country European Statistical System network (ESSnet) projects, intended to facilitate the integration of geographical and statistical information: GEOSTAT projects 1 to 4. The first project in the series addressed methodological questions regarding the dissemination of grid data. GEOSTAT 4, which is ongoing, includes a study of the implementation conditions for the Global Statistical Geospatial Framework at the European level. It is also aimed at defining a quality reference framework for the production and dissemination of highly localized statistics.
- Eurostat funds national initiatives on geographical integration through a specific system of grants.
- Eurostat funds the activities of the European Forum for Geography and Statistics. The Forum also brings together members of the two communities, within a voluntary organization. For several years, its annual conference has been an occasion for rich discussions that now involve more than 200 participants.

5. INSEE has participated in all the GEOSTAT projects (except GEOSTAT 3). As part of the ongoing GEOSTAT 4, INSEE is actively contributing to work on convergence between the Global Statistical Geospatial Framework and the models for describing statistical processes produced by the Economic Commission for Europe, including the Generic Statistical Business Process Model, the Generic Activity Model for Statistical Organizations, the Common Statistical Production Architecture and the Generic Statistical Information Model.

6. INSEE produces statistics and carries out studies based on the data it produces. For the entire community to take advantage of this difference between INSEE and other national statistical institutes, Eurostat commissioned and co-financed the development by INSEE of a spatial statistics handbook. This work was also supported by the European Forum for Geography and Statistics. Unlike other initiatives, which concern the conditions for setting up or rolling out a fine-grained georeferenced statistical information system, the handbook is written on the assumption that such a system exists. The handbook's 14 chapters describe various available processing methods for spatial statistics, designed to enhance or facilitate production, dissemination or statistical analysis operations. Aimed specifically at national statistical institutes, the handbook is focused on issues rarely addressed in academic equivalents, such as spatial sampling or the confidentiality of statistics and geographical data. The handbook is distributed free of charge in French and English on the websites of INSEE, Eurostat and the European Forum for Geography and Statistics and has been downloaded more than 10,000 times in two years.

7. Lastly, INSEE and IGN jointly organized the conference of the European Forum for Geography and Statistics held in Paris in November 2016. In addition to avenues for collaboration between the two communities, the conference agenda was focused on spatial analysis and academic outreach.

II. Geography and statistics at the national level

8. Historically, INSEE has met its needs internally, by managing its own geographical information system and integrating it with the statistical information system. This type of structure is considered "in-house" for the purpose of the GEOSTAT 2 project. The new but already fruitful collaboration with IGN has not called into question the fundamental principles of this set-up. The current approach of IGN, which is in line with the principles of the Global Statistical Geospatial Framework, opens up the possibility, in the medium-term, of a division of roles more suited to the specific mission of each institute.

A. Current organization at INSEE

1. Until 1999

9. The population census was the main statistical output reliant on geographical information. In France, without a localized national population register, the census was taken using the "traditional" method. Citizens filled in a self-administered questionnaire delivered by census takers.

10. The work of the census takers was organized by assigning each of them a collection area with a number of housing units that constituted a suitable workload and geographical limits that were easily identifiable on the ground. Until 1999, these collection areas were hand-drawn on IGN paper maps in INSEE regional offices. In the biggest cities, the collection areas were combined into slightly larger areas (with approximately 2,000 residents), referred to as aggregated units for statistical information (known by the French acronym of “IRIS”), for the purpose of disseminating the statistical results of the census.

11. Meticulous coding of the collection and dissemination areas meant that there was no reliance on an accurate geographical information system, which in any case did not exist.

2. 1999 to 2016

12. The French census was taken once per decade until 1999. In order to reduce census production and dissemination times, in 2004, INSEE moved to a system of annual census surveys. Each year, 8 per cent of residential addresses in large municipalities with more than 10,000 inhabitants were randomly included in the census, while one fifth of small municipalities were surveyed in their entirety. The annual census itself was based on consecutive annual surveys taken over a rolling five-year period and so covered 40 per cent of the residential addresses in large municipalities and 100 per cent of those in small municipalities. The questionnaire remained self-administered (paper or online as chosen by the respondent).

13. In small municipalities, the move to annual census surveys did not fundamentally change the need for geographical information. Maps were still used to define collection areas. In large municipalities, the change was more profound. A register of residential addresses was required. With no external supplier able to meet the needs of INSEE within the time frame for implementation of the new census project, INSEE created its own address directory (register of localized buildings, or RIL) and established the conditions for its annual maintenance as part of a partnership with municipal authorities. The addresses in the register were described semantically but also geocoded to facilitate their identification on the ground and for automatic allocation to areas for dissemination purposes. In addition to the address directory, INSEE managed the geometric borders of municipalities and the dissemination areas for large municipalities.

14. Subsequently, the register was used for the georeferencing of administrative files to meet other public statistics needs. Since there was no interoperable address identifier or common standard for address description, the addresses of the statistical units in the file to be georeferenced were systematically compared to the addresses in the register of localized buildings. The quality of the result thus depended on the quality of the register in terms of freshness and exhaustiveness, the quality of the pairing engine and of the address inputs in the file to be georeferenced and the general quality of addressing in France. On that last point, there are many residences in France with unnumbered addresses. This occurs in rural hamlets, for example. In order to ensure the statistical quality of the disseminated results despite this fact, a 15-member team was tasked with manually processing cases of poor-quality pairing, regardless of the reason for the pairing error.

3. 2016–present

15. In 1999, the French Government commissioned IGN to design and produce a large-scale reference system (RGE®) providing an accurate, comprehensive and consistent description of the national territory and land use within it. Available since 2006, the reference system consists of five components (orthophotography, topography, altimetry, parcel and address), which provide fully superimposable geographical information. Since 1 January 2011, the reference system has been distributed for the cost price of reproduction and dissemination of the relevant data, as long as the data are for public service use with no industrial or commercial purpose.

16. The fact that the reference system was first disseminated in 2006 and especially the way it was disseminated until 2011 were barriers to its use for INSEE activities. The first major overhaul of the rolling census from 2013 was an opportunity to revisit the partnership between INSEE and IGN.

17. The first observation made by INSEE was that its in-house management of geographical data in the register of localized buildings and the related boundary directory

led to topographical inconsistencies when information from the register of localized buildings was displayed on RGE® backgrounds. For example, addresses from the register for Saint-Malo (in Brittany) were located in the sea according to RGE®. In order to address these inconsistencies, INSEE and IGN formed an initial partnership for the purpose of aligning geographical information from INSEE, including the position of addresses and the boundaries of IRIS areas, with the information from RGE®. Subsequently, to maintain the consistency achieved in their geographical reference information, INSEE and IGN signed two agreements.

18. The first agreement gave IGN full responsibility for managing the geographical boundaries required by INSEE, including the IRIS boundaries. The second established the following conditions for mutually advantageous exchanges in the future: IGN provides the map layers and cartographic services required by INSEE for it to locate new residential addresses identified in the context of its partnership with the municipal authorities in a manner consistent with RGE®; IGN also incorporates new roads identified by INSEE within 10 days. In return, INSEE provides IGN with its store of localized residential addresses and the flow of new addresses.

19. This way of working remains primarily in-house. INSEE continues to manage its own directory of addresses and incorporate new addresses, although these are identified and located in a manner consistent with RGE®. This partial outsourcing is due to the fact that, in 2013, the RGE® address component did not meet the technical specifications set by INSEE for freshness, exhaustiveness and interoperability of address identifiers and the ability to separate out residential addresses, for example from those used exclusively for economic activities.

20. The national address database (BAN) project came into being at the same time. The database results from an unprecedented partnership between IGN, Etalab, the La Poste (post office) group, the Directorate General of Public Finances and OpenStreetMap France. The aim of this free collaborative database is to reference the geographical position of all addresses located in French territory. Finally produced in 2018 by the interministerial directorate responsible for digital technology, IGN and La Poste, it is continuously updated by local authorities, the emergency services, institutions such as the Directorate General of Public Finances, INSEE and IGN, companies and individual citizens. IGN provides a collaborative portal for municipal authorities, allowing their data to be directly integrated in the national address database. In parallel, mass data imports not only from INSEE but also from La Poste and the Directorate General of Public Finances contribute to the maintenance of the database. Since 1 January 2020, the database has been distributed under an open licence. The licence was designed by the State to facilitate and encourage the reuse of freely available public data.

B. Future prospects

1. Authoritative geographical data and common georeferencing procedures

21. The aim of principle 1 of the Global Statistical Geospatial Framework is to ensure that statistical results observed locally from different sources genuinely reveal the underlying social or economic phenomena and not artefacts linked to differences in the geographical data used for each source or different georeferencing methods. It therefore promotes the use of authoritative geographical data and of geographical integration methods common to all statistical production processes.

22. In line with this principle, a study on sovereign geographical data was launched in France in 2018. Defining "sovereign" geographical data as those used directly as a decision-making aid by the public authorities, a parliamentary report[5] submitted to the Government in 2018 contained various recommendations on such data and the methods for their production and dissemination, possible synergies between them and governance processes that would enable better coordination between stakeholders.

23. The author of the parliamentary report supported the aim of a national address database, while recognizing its limits, and went so far as to recommend investigating, in

[5] V. Faure-Muntian, *Les données géographiques souveraines* (Government report, 2018).

consultation with organizations representing the municipalities, the possibility of introducing a general obligation to produce addressing plans and to allow appeals against the allocation of addresses. All this will tend to harmonize and improve the efficiency of statistical production processes. Therefore INSEE, working closely with IGN, is attentively following the development of the national address database and actively contributing to it.

24. In return, these possibilities have led INSEE to add a major overhaul of geographical data management, based on the national address database, to its own programme of work. This could eventually concern all statistical production, including the census, the georeferencing of administrative files and company directories. In accordance with the principles of the Global Statistical Geospatial Framework, the objective is for geographical data to be a clearly identified component of statistical business processes within the meaning of the Generic Statistical Business Process Model, based on electronic infrastructure adequate to the aims of streamlining and efficiency, including through the development of application programming interfaces in which geographical or statistical references can be queried directly rather than through local copies of the reference information.

2. Further opportunities for collaboration

25. Beyond infrastructure matters, additional avenues for cooperation have recently emerged and could be made systematic.

26. IGN operates the French national territorial portal (http://geoportail.gouv.fr). In this capacity, it has made available online a cartographic representation of the grid data on population and income produced by INSEE (see figure 1). For the moment, this output is limited to a single statistical source. In future, it should be expanded to include the 2021 census data, in line with the European regulation on the census. Data on employment and businesses could also be disseminated in grid form.

27. This dissemination of statistical data in map form will be made easier by the fact that both institutes apply the Resource Description Framework (RDF) to disseminate their output. The Framework is used to produce a formal description of sources and their metadata, to enable automatic processing and a level of interoperability.

28. INSEE is currently finalizing the conversion of the territorial names under its responsibility into the format of the Framework. INSEE and IGN are also in contact to work on the equivalent formatting process for the geometric equivalents of territorial names. When this work is completed, searches for statistical and geographical data will become more dynamic and efficient, in line with the principles of linked open data.

III. Conclusion

29. Owing to advances in the geographical data available, INSEE has recently moved from the exclusively in-house management of such data towards a closer relationship with IGN, without fundamentally changing its way of working. Key projects to create an address directory initially prompted INSEE to perform an in-depth review of its geographical management and to embark on an approach involving closer cooperation. This infrastructure creation enterprise will involve further opportunities for collaboration to the mutual advantage of the two communities, including on the dissemination of highly localized statistical data.

Figure

Online provision by IGN on the French national territorial portal of INSEE grid data (proportion of poor households)