Implementing the GSGF at Insee

Vincent Loonis
Insee

UN - EG ISGI, Manchester October 8th, 2019
Insee has a longstanding history in the management of Geospatial and Statistical information.

Until recently, this management was *in-house* and essentially aimed at meeting the census’s organisational needs.

It is not suitable for feeding the growing demand on small area statistics.

In order to achieve its "2025-vision" of a fully geocoded statistical information system, Insee has to reshape its management of geospatial information.

The new system has to be more compliant with the GSGF’s principles, and will more or less be that of a *data broker*.

There is a long way to go...
The French statistical information system is not a register-based system.
- There are no inter-connected address, building, dwelling or population registers sharing common IDs.
- The existing statistical or administrative files are not directly connected to the geospatial information.

The census is a traditional census.

The geocoding of statistical or administrative files relies on address matching with a list of addresses managed by Insee.

Due to the quality of the addresses, the quality of the matching may vary across files and locations (rural or urban areas).

There is a 15-people team devoted to manage the mismatches.
The new Insee system will rely on a national address register that is currently under construction.

It will use a common tool for the matching of addresses.

Insee is trying to be part of the governance to be sure that the new tools will meet its needs.

To manage the mismatches, whose number will probably not decrease, Insee will nevertheless have to: set up a new *in-house* automatic geocoding system relying on pseudo-IDs and cadastral information, keep a reduced devoted team for the remaining mismatches, set up a quality management of the geocoding.

Insee is fostering the construction of new registers (building or dwelling) and the use of commons IDs at least within the official statistical world.
Insee is trying to build a data management environment that meets Geostat 3 recommandations, while taking into account the French specificities (no registers, no common IDs), being flexible enough to be compatible with any innovations (new building register, new pseudo Ids).
Principle 3: Common geographies

The same grid system for all databases.

**Figure:** Splitting squares until meeting a given threshold (number of observations) that depends on the database. All the released information is the true one, but the squares are of various sizes.

**Figure:** Disseminating information on same-size cells (200 m or 1 km) but with imputed values when the number of observations is below the threshold.
**Figure:** Using only the European reference system to avoid geo differencing while disseminating with two different systems (National or European)
Principle 4: Interoperability

Figure: Insee is responsible for both official and statistical territorial classifications. Insee is about to disseminate according to the 4-star standard. The 5-standard is now reachable.
Insee released last June a 200-meter-grid database with data on dwellings, incomes, and population by sex.

Insee aims at disseminating other data as soon as possible.

The data are available, for free, as files as well as on our NMA geo-portal.

Insee has been very careful about data protection and privacy issues.

Insee has fostered the use of geocoded statistical information with the dissemination of a Handbook of spatial analysis, supported by EFGS and Eurostat.