28th Meeting of the Wiesbaden Group on Business Registers   - International Roundtable on Business Survey Frames
The Hague, The Netherlands, 2 – 6 October 2023   *Masahiko Yagi, Ilaria Di Matteo, Zhiyuan Qian, Pedro Farinas, Mark Iliffe (UNSD)*   *and Hank Hermans (CBS) and Claudio Stenner (IBGE)*   *Session No. 5*   Workshop SBR Maturity Model (SMM)
Statistical Business Registers and Geospatial Information

Abstract

An increasing number of countries are integrating geospatial information into Statistical Business Registers (SBRs). The integration of geospatial information in the SBRs represents an important step toward the development of an integrated statistical geospatial infrastructure. Recognizing the importance of this integration, the United Nations Statistical Commission, at its 53rd session in March 2022, requested to the Committee of Experts on Business and Trade Statistics (UNCEBTS) to develop guidelines to assist countries to integrate geospatial information into SBRs.

The UNCEBTS Task team on Statistical Business Registers, in collaboration with the UN Expert Group on the Integration of Statistical and Geospatial Information (EG-ISGI), is developing a report describing what it means to integrate geospatial information into SBRs, what the benefits of such integration are, how to integrate geospatial information in practice and how to maintain this integration in a sustainable manner. The report will build on existing country practices. The close collaboration with the EG-ISGI will ensure that the guidelines are fully aligned with the Global Statistical Geospatial Framework (GSGF). The report will be presented to the UN Statistical Commission in March 2024. The paper will elaborate on the work programme of the Task team on SBRs and the EG-ISGI.

Introduction

1. Statistical Business Registers are a fundamental element of the statistical infrastructure for economic statistics as they provide the population universe for the construction of sampling surveys of economic activities, support the compilation of business demography statistics and support the integration of data from different sources. Recognizing the potential of geospatial information for SBRs, the United Nations Statistical Commission, at its 53rd session in March 2022, requested the Committee of Experts on Business and Trade Statistics (UNCEBTS) to develop guidelines to assist countries integrate geospatial information into SBRs[1], through establishing the UNCEBTS Task team on Statistical Business Registers.

2. To be fully effective, all SBRs should be linked to a geographic location, commonly known as 'geocoding'. Traditionally, SBRs were geocoded by the address of the business unit. Yet, prevailing practice by an increasing number of countries 'geocodes' SBRs also through an x- and y-coordinate. By geocoding with x- and y-coordinates the production, measurement, monitoring and dissemination of SBRs becomes easier, enabling National Statistical Offices (NSOs) and other SBR-producing agencies to provide SBRs that meet a wider range of needs and applications, in effect, geospatially enabling them.

3. The geospatial enablement of SBRs can be informed by the Global Statistical Geospatial Framework (GSGF)[2]. The GSGF facilitates the integration of statistical and geospatial information. A Framework for the world, the GSGF enables a range of data to be integrated from both statistical and geospatial communities and, through the application of its five Principles and supporting key elements, permits the production of harmonised and standardised geospatially enabled statistical data. The resulting data can then be integrated with statistical, geospatial, and other information to inform and facilitate data-driven and evidence-based decision-making to support local, sub-national, national, regional, and global development priorities and agendas. The GSGF was developed by the UN Expert Group on the Integration of Statistical and Geospatial Information[3] (EG-ISGI), as an Expert Group that reports to both the UNSC and United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), as the apex intergovernmental communities for statistical and geospatial information respectively. The GSGF was first endorsed by UN-GGIM, by its decision 9/106[4], a decision endorsed by the UNSC by its decision 51/123.

4. Against this background, the UNCEBTS Task team on Statistical Business Registers is closely collaborating with EG-ISGI to develop a report describing what it means to integrate geospatial information into SBRs and respond to UNSC decision 52/107 [5]. This report will detail the benefits arising from implementing the GSGF for SBRs, highlight how to integrate geospatial information in practice and how to maintain this integration sustainably. Moreover, the report will highlight existing good national practices. The report will be presented to the UN Statistical Commission in March 2024. The paper will elaborate on the work programme of the Task Team on SBRs and the EG-ISGI.

---

[1] See decision 53/119 of E/2022/24 E/CN.3/2022/41 available at:
 https://unstats.un.org/UNSDWebsite/statcom/session_53/documents/2022-41-FinalReport-E.pdf
[2] https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf
[3] https://ggim.un.org/UNGGIM-expert-group/
[4] E/2020/46-E/C.20/2019/19
[5] https://unstats.un.org/UNSDWebsite/statcom/session_52/documents/2021-30-FinalReport-E.pdf

What is geospatial information and its links with SBRs

5.   As described in the United Nations Integrated Geospatial Information Framework (UN-IGIF)[6], geospatial information describes the physical location of geographic features and their relationship to other features and associated statistical information. It is presented in many forms such as maps, satellite imagery and aerial photography.

Figure 1 The Global Statistical Geospatial Framework
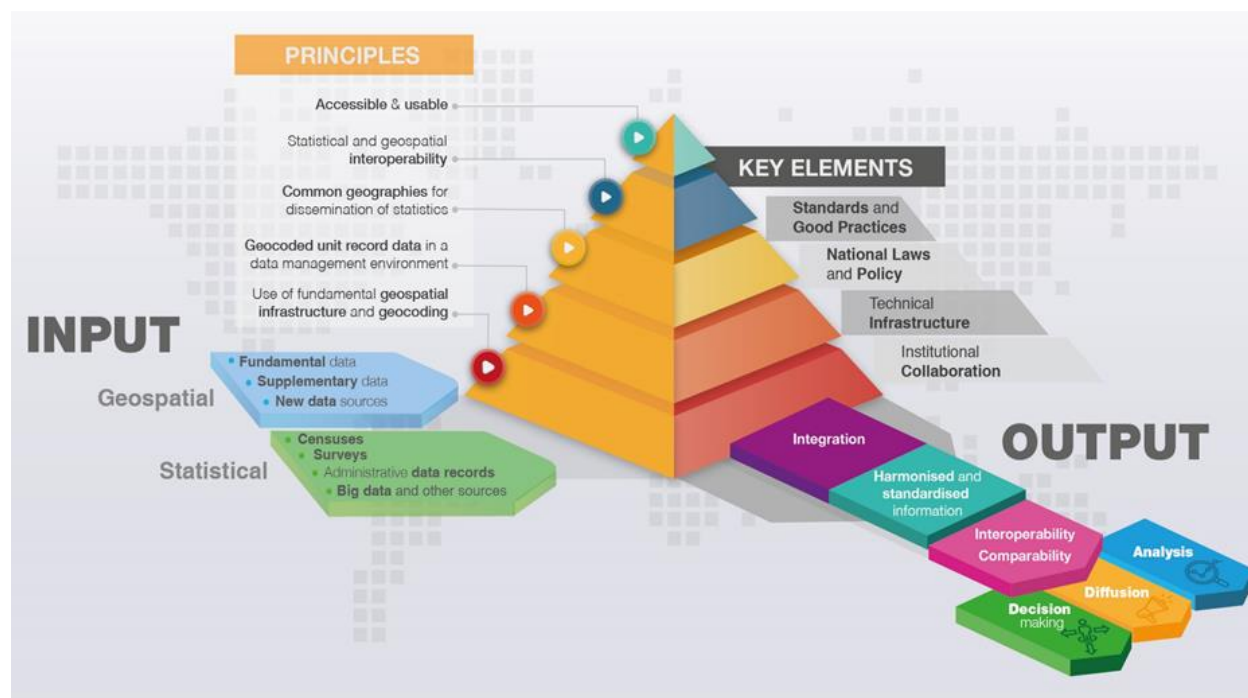
6.   The Global Statistical Geospatial Framework (GSGF) provides the common framework that links between statistical and geospatial professional domains, between NSOs and National Geospatial Information Agencies (NGIAs)[7], crossing the divide between statistical and geospatial standards, methods, workflows, and tools. The GSGF begins with critical inputs of Fundamental Geospatial Data, and supplements these with other geospatial data sources as necessary, which can be provided by NGIAs and the broader geospatial community. It primarily serves to geospatially enable traditional and authoritative statistical data, and increasingly data from administrative and other sources, much of which comes from NSOs and administrative data custodians within the broader National Statistical System (NSS). Each of

---

[6] https://ggim.un.org/IGIF/documents/Part_1_UN-IGIF_Overarching_Strategy_Second_Edition_27Feb2023.pdf
[7] NGIA is used as an encompassing term to cover National Mapping, National Cartographic, National Geospatial Information Agencies and Authorities.

these inputs could be augmented by complementary datasets, such as those derived from crowdsourcing, new technology or sources (e.g. Big Data)[8].

7.   The GSGF contains five principles that describe the broad processes by which a range of geospatial and statistical infrastructures and processes are applied to input data to enable integration, namely:

1) Use of fundamental geospatial infrastructure and geocoding;

2) Geocoded unit record data in a data management environment;

3) Common geographies for the dissemination of statistics;

4) Statistical and geospatial interoperability; and,

5) Accessible and usable geospatially enabled statistics.

8.   All these principles are relevant and applicable for SBRs although some more than others. An essential initial element, however, is the use of fundamental geospatial infrastructure and geocoding in the SBRs (Principle 1), geocoded unit record data in a data management environment (Principle 2), and the use of common geographies (Principle 3). The description of these principles is presented in boxes 1, 2, and 3 below.

Box 1: Principle 1: Use of fundamental geospatial infrastructure and geocoding

Principle 1 specifies the adoption of a common and consistent approach to place each statistical unit of a dataset in time and space, using fundamental geospatial infrastructure.

The goal of Principle 1 is to obtain high quality, standardized location references (such as physical addresses, property or building identifiers, or other location descriptions), in order to assign accurate coordinates, and/or a small geographic area or standard grid reference, to each statistical unit at the microdata/unit record level. In addition, time and date-stamping these locations clearly places the unit both in time and in space, due to the strong statistical requirement for establishing data in a time series. Preferably, location is recorded through direct or indirect capture of x- and y-coordinates. Where this level of precision is not possible using current geospatial and statistical infrastructure within a country, adaptations using more general location descriptions and/or larger geographies will be necessary.

The process of obtaining locations and geocodes should use relevant, fundamental geospatial data, and support system capabilities from a NSDI or other nationally agreed sources. Ideally, a country should have a geo-referenced address, building register, land parcel, and/or place name as part of its statistical-geospatial infrastructure. If such a register is not available, countries are urged to test the implementation of alternative point-based referencing for unit record data .

Implementation of Principle 1 achieves the following objectives:

- Address, property, building, and location information are accurate and consistent, meeting country-level agreed standards and good practices;
- Geocoding results are as accurate and consistent as possible using common approaches or systems; and,
- Any geocoding issues are consistently managed through application of standardized approaches.

---

[8] GSGF

Box 2: Principle 2: Geocoded unit record data in a data management environment

Principle 2 supports the process of linking or storing high-precision geographic references (i.e. geocodes – coordinates, small geographic area codes, or linked-data identifiers) to each microdata/statistical unit record. This is often referred to as geospatially enabling data, and must occur within a secure, standards-based data management environment. This process applies the address coding infrastructure and fundamental data from Principle 1.

The goal of Principle 2 is to allow all statistical unit records to be linked to a location, wherever it is possible to do so. This will then enable integration of data from a wide variety of sources, such as other socioeconomic statistical data, administrative data, and geospatial information about the built and natural environment. The incorporation of these data, using geospatial processing, can then deliver new, geospatially enabled statistical variables for analysis. Principle 2 also enables flexible application of any geographic content when preparing data for release and analysis. This includes supporting future aggregation of statistical data into new geographical units or adapting to changes to existing geographies over time.

Principle 2 includes the use of data management tools, techniques, standards and good practices to facilitate the linking and management of geocodes within statistical datasets. This also serves to ensure that privacy and confidentiality requirements are correctly managed for the released data.

Implementation of Principle 2 achieves the following objectives:

- All statistical microdata is geospatially enabled for flexible use in analysis, visualization, dissemination and statistical data integration processes;
- Aggregation of data for larger geographies is simplified through storage of a unique identifier or code for a small area geography or standard grid cell for each unit record;
- Adaptation to changes to existing geographies or to allow compilation of data for new geographies is enabled;
- Data can be effectively managed, including the protection of privacy and confidentiality;
- Clear data maintenance and custodianship roles are defined; and,
- Geocoded information and metadata are consistent, interpretable and systematically maintained.

Box 3: Principle 3: Common geographies for dissemination of statistics

Principle 3 applies geography as a tool for integrating data. It uses a common and agreed set of geographies for the display, storage, reporting, and analysis of social, economic and environmental comparisons across statistical datasets from different sources. Principle 3 establishes the fundamental importance of balancing existing statistical and administrative geographies with other geographic referencing systems, such as grids, as a basis for establishing common geographies across datasets.

The goal of Principle 3 is to support the provision of a common set of geographies that ensure the consistent geospatial aggregation and dissemination of statistical data, irrespective of whether they are in gridded or administrative boundaries. Data is uniformly allocated to smaller administrative segments or statistical units (such as mesh blocks) that are divided according to political, property or topological subdivisions, or consistently assigned to differently sized grid units (i.e. squares or pixels). Furthermore, Principle 3 also allows for the translation and mapping of statistical information between gridded and administrative boundaries.

Implementation of Principle 3 achieves the following objectives:

- Data from different sources can be integrated using a common geography;
- The visualisation, analysis and interpretation of statistical and geographic information is simplified;
- Metadata supports data aggregation, integration and use;
- Identification and application of aggregation and disaggregation methods will enhance data quality and the assessment, consistency and increased use of data; and,
- The conversion of data between geographies is supported, through standard conversion mechanisms (e.g. through correspondences)

9.   From these three principles it is evident that the inclusion of geocoded information in the SBR is crucial to the development of a geographical information system integrated in the statistical production process. Some other advantages in geospatially enabling SBRs is that to have more precise sampling strategies when links to spatial areas are important; improve integration and linkage of business activities with environmental and social information in a spatially explicit manner; and improved methods for small area estimation. In addition, interoperable systems and standard metadata (Principle 4), and technologies for the dissemination and visualization of geospatially integrated statistical data (Principle 5).

10.   Ultimately, the geospatial enablement of SBRs expands the types of analysis that can be done and allow for further insights. "The association of geography and statistics has the potential to generate information far beyond the simple representation of data on a map. Linking geo-referenced and numerical statistics in spatial analysis has the potential to reveal relationships and phenomena which are difficult to discover by analyzing statistical databases alone."[9]

---

[9]Merging statistics and geospatial information - experiences and observations from national statistical authorities, 2012-2015 projects, 2019 edition Eurostat. Available at
https://ec.europa.eu/eurostat/documents/7870049/9809882/KS-FT-19-004-EN-N.pdf/aec76d11-967e-4d67-aad2-49c52b264748?t=1559812901000

11.   The report on geospatial information and SBRs will make specific reference to what the Principles of the GSGF mean in practical terms for the SBRs.

How to geospatially enable SBRs

12.   In general, the inclusion of geocoded information in the SBR should be viewed as part of a broader approach to the integration of geospatial information and statistics which is provided by the GSGF. Elements of this framework include, for example, the institutional framework for this integration which brings together NSOs and NGIAs. The guidelines on the integration of geocoded information in SBRs will focus on specific aspects related to SBRs and, whenever relevant, reference will be made to the broader framework including the institutional arrangements.

13.   The guidelines for introducing geospatial information in the SBRs is expected to be practical in nature and address some basic questions such as: how to convert addresses into geocoded x- and y-coordinates; how to maintain this information; how to collaborate with NGIAs to maintain and update this information; how to address specific implementation issues in assigning geocodes, etc.

Way forward

14.   The Task team on Statistical Business Registers of the UNCEBTS and the EG-ISGI of the UN GGIM are working together on a report on the integration of geospatial information in the SBRs to be presented to the UNSC in 2024. The report will build on the country practices of several countries that have successfully integrated geospatial information and statistics, such as the experience of INEGI in the development-Mexico of DENUE, IBGE- Brazil, and several other countries that have pioneered this integration.