



Economic Commission for Europe

Conference of European Statisticians

Seventieth plenary session

Geneva, 20-22 June 2022

Item 9 of the provisional agenda

Collaboration with private data providers

**National Institute of Statistics and Geography's (INEGI)
collaboration with private data providers: Lessons,
challenges and opportunities**

Prepared by Mexico

Summary

This document describes the experience of Mexico's National Institute of Statistics and Geography (INEGI) in collaborating with private data providers.

The document is presented to the Conference of European Statisticians' session on "Collaboration with private data providers" for discussion.



I. Overview

1. In recent decades, and even more so in the past few years, the relationships between people and data have changed dramatically – in quantity, quality, as in complexity. These quantum-like shifts have entailed both advantages and challenges which, albeit of different natures, are equally exciting and daunting for those involved. For citizens/consumers, challenges are mainly related to privacy and use/administration of time, among others. For private entities/corporations, it's mostly a matter of intellectual and commercial property, and a constant dilemma between profit (doing right by shareholders) and benefit (doing right by consumers). For public data providers, particularly national statistical offices, it is a matter of relevance: how to still provide accurate and quality data to users, while keeping their privacy, and keeping up with increasingly diverse data ecosystem which outcompetes them in pizzazz and opportunity. In other words, for NSOs, it's a matter of sheer survival.
2. In this ever-challenging context, Mexico's National Institute of Statistics and Geography (INEGI), has certain particularities. One, it is both the country's NSO, and its national cartographic/geographic entity – a feature which grants the potential for an integrated analysis of statistical data in a spatial context. Taken to its furthest, this allows for a greatly enhanced design, monitoring, and evaluation, of public policies and indicators, including SDG and other international frameworks.
3. Another structural peculiarity of INEGI is its dual role as an NSO (official producer of information) and the coordinator of Mexico's National System of Statistical and Geographic Information (SNIEG, for its Spanish acronym) – whereby it serves as technical secretariat to several collegiate bodies, involving all information producers within the State (and other stakeholders, including subnational governments, academia and the private sector), so that everyone participates in the design and approval of standards, methodologies, instruments (such as census and surveys) and ultimately, official data, which can be widely used with the benefit of quality, opportunity, comparability and, ultimately, TRUST.
4. These unique features are enshrined within our Constitution. Consequently, our legal framework grants INEGI relatively ample attributions to enquire both individuals and corporations for information, while safeguarding their privacy and anonymity through clear and strict guidelines on quality and confidentiality, based on international standards, as well as on our General Law on Data Protection. All these factors result in a considerable degree of trust in INEGI, reflected in different studies and surveys. This trust, together with the importance of the information it produces for decision-making, places the institution in a favorable position to request privately generated or held data.
5. Regarding collaborations with private data providers, a cross-cutting element identified is the important role of public institutions. Mexico has a solid and broad public institutional structure, with powers and attributions that imply direct contact or regulatory relationships with holders of private data. It is also possible that public institutions generate private data, e.g., social security, tax and revenue institutions, the central bank, and other credit and development financial institutions. Coordination and communication between certain public institutions during different stages of a project have reported benefits or revealed future opportunities, e.g., proper coordination between public agencies to consolidate purchases can result in lower costs or in avoiding burdens for data holders, and, therefore, greater willingness to share their information. Another set of relevant institutions are the regulatory bodies, like the ones for telecommunications, financial or economic competition; depending on their faculties, these institutions can act as trusted intermediaries in access to data or to give an opinion on whether a project with private parties may harm competition.
6. The growing demand for more timely and frequent information, with greater disaggregation, together with other features linked to direct information needs from new users, has posed a significant challenge to statistical offices, together with increasing budgetary restrictions or requirements – to satisfy growing data needs – coupled with the understanding of their enduring and important role in guaranteeing quality, comparable and strictly anonymous data. Given that our legal framework does not explicitly regulate access

to third-party data held by the system's informants, projects and initiatives undertaken by INEGI have required the signing of specific and unique collaboration agreements.

7. Regarding the usability and quality of private data, we have identified it can be used as support material for processes that INEGI carries out regularly, e.g., to identify new settlements, to update cartography, to improve the timeliness or disaggregation of key indicators, including those belonging to international frameworks such as the SDGs. Private data can also be used to generate new statistical information, such as detailed car sales, private consumption, mobility, or wage dispersion. In any case, there is a fundamental element that statistical offices must consider: quality of information. Privately held data is, generally, structured under the logic of the (commercial) activity that the informant develops, so its use for official statistics requires structuring for a proper statistical analysis and use that allows obtaining the variables of interest. Therefore, a number of specific actions can be implemented for each project, among which, the most relevant are: i) in the publication of the information, it should be agreed that the data can be updated; ii) the implementation of mechanisms to verify certain consistency in the information whose results are shared with the data providers; iii) the design of technical sheets and pseudocodes that are shared with those in charge of processing the information, and iv) building synthetic databases that simulate those of the servers of the data holders.

8. New data sources, including those from the private sector, have shown great potential to augment and complement official statistics. In many cases, they allow the delivery of more timely and relevant statistics, corresponding to rapidly changing socioeconomic phenomena and dynamic changes in the needs of data users. COVID-19 reinforced the importance and need to have new data sources and alternatives for the information collection (such as experimental statistics), considering the difficulties encountered in collecting information in households and businesses. Challenges and lessons learned vary according to specific projects or thematic areas, as described in the following section.

II. Surging cloud of private data providers: delving into new waters and the resulting challenges

9. Access to private data to INEGI has been granted for different objectives and in different contexts, e.g., by particular initiatives or projects through requests from INEGI, while others have been triggered by request from private institutions or civil organizations who wanted their data to be visible and used. These examples show there is both need, openness, and possibility, but also that demand-driven windows are few, far-between and unable to respond to real information needs unless subject to coordination and systematization, standards and terms.

10. The knowledge obtained through these projects can be vast, and open our perspectives to new possibilities, analyses, and synergies. They also make us identify new challenges and needs to adapt and cope – hopefully together.

11. Currently, INEGI is undertaking the following projects to access private data:

A. Data collected by the banking system

12. The Project consists of the transfer of statistics calculated from micro aggregations of bank card transactions. This will enable INEGI to publish timely monthly information with different levels of geographical breakdown, as well as for the different types of channels related to private consumption such as cash withdrawals, purchases with a card present (POS) and purchases with a card not present (Digital). Additionally, it will enable the generation of information based on administrative records on payrolls, which will contribute to greater knowledge of the labour market in Mexico.

13. Bilateral collaboration agreements were signed with three of the main financial institutions in Mexico, which together concentrate just over half of the market. The COVID pandemic required the need for new sources of information, and this was internalized by these banking institutions.

14. The project is currently in the development phase, in which quality of information issues are being addressed to ensure a homogeneous treatment of raw data within banking institutions.

15. Some of the benefits that banks can obtain from this project come from the publication of the information itself or from its processing. The most relevant benefits identified are the following:

(a) Visibility of its social responsibility by contributing to an institution with a high reputation and reliability such as INEGI, whose information is disseminated and used in practically any economic and social context in the country.

(b) Better knowledge about their markets. In Mexico, the public information on banking activity is carried out by regulatory institutions, such as the Central Bank. However, this information has important restrictions, such as geographic disaggregation or the sociodemographic characterization of the population. Given that the project does contemplate these edges, future analyses that the banks can carry out will be favoured.

(c) Identification of improvements in internal processes. The data requests we have made about aggregations, linking of its information, or verifying statistics, have contributed to a greater understanding of our partners and of the process itself, or to the identification of potential internal improvements.

(d) Better knowledge about the potential of its data. On a regular basis, the main banking institutions publish reports or documents on various aspects of the country's economic activity, generally based on their own data and taking as a reference various economic and social indicator generated by INEGI. The publication of the information resulting from this project, in addition to its usability as is, will serve to complement or expand the possibilities of indicators on the economic and social life of the country, such as data on employment, consumption, sales, etc. In this sense, the publication would serve the institutions to evaluate the scope and limitations of their data in a broader context.

(e) In addition to the benefits of the publication itself, we have identified during the project elements that, without constituting a payment (monetary or in kind), have been developed by INEGI and that have been of interest to the banks. For example, based on simulated information and information that has been shared with us, we develop visual elements to represent the data and generate new insights. These innovations have been attractive to data holders and sharing them has favoured the flow of communication and perception of a tangible and immediate benefit for the banks.

B. Data from the automotive industry, including light and heavy vehicles

16. The signing of a collaboration agreement in 2018 with the Mexican Association of the Automotive Industry (AMIA, for its Spanish acronym) and its 21 affiliated companies made it possible for INEGI to publish the *Administrative Registry of the Light Vehicle Automotive Industry* (RAIAVL) monthly. The positive results derived from the Registry encouraged the nine companies affiliated within the National Association of Producers of Buses, Trucks and Tractor Trucks (ANPACT) to request and implement a similar registry with INEGI in 2020. For the first time, data was provided by, and disseminated at, the company level. To achieve this, it was necessary to attain the authorization from each of the companies to share their data, while also giving it the nuance and treatment of primary/raw data, which would not have any type of statistical treatment, nor be subject to the regulations and filters of the rest of the official statistical programs. Before publishing the figures of these administrative records, the Federal Economic Competition Commission (COFECE), the authority that regulates competition in free market activities, was consulted on the effects of disseminating this information to avoid any negative effects on competitiveness and guarantee that the data provided would not represent a risk to competition between companies in these sectors.

17. The success of this collaboration rested on the following factors:

(a) Private information providers identify the data provided to INEGI as a priority for decision-making in different areas such as sales strategies, production, export, etc. Being

a fundamental input for the development of their market analysis activities, companies comply with the requested information in a timely manner.

(b) Development of a specific digital platform to exchange data in a secure simple way.

(c) Signed collaboration agreements: to guarantee the flow of information to INEGI, each of the affiliated companies committed to provide information in a monthly and timely manner and on established dates. Likewise, they granted their consent for INEGI to disseminate their information at the brand and company level. This offered data providers greater certainty and clarity on how their data would be used and disseminated, with the same rules for all participants.

18. This compromise represented both a paradigm shift in INEGI, assuring access, continuity, and interoperability between public and private sources, for the benefit of both partners, and a challenge. The disadvantage of this type of information is that updates by companies may be irregular and inconsistent (in both frequency and quality), often showing significant changes. The quality of the data provided by the automotive companies can be ensured through validations carried out by the system, as the informants capture their information. Also, a comparison is provided where the companies visualize the accumulated behaviour of the data previously provided. Likewise, errors in the provision of information can be corrected by companies in subsequent updates. Thus, as data is shared, processed, and presented, both partners understand its importance, and can improve their gathering and processing in a progressive manner.

C. The multiplying effect of looking beyond, together

19. Participating in international groups and initiatives has frequently yielded interesting and positive results. To name but a few examples:

(a) Mexico's participation within the United Nations Group of Experts on Geospatial Information Management (UN-GGIM) and the Group on Earth Observations (GEO) has been a fruitful experience, in terms of technical, economic, and diplomatic exchange, where we have been providers, brokers and recipients of cooperation agreements. Together with Chile, and with the support of the UN Regional Economic Commission for Latin America and the Caribbean, Mexico (through INEGI and the Mexican International Cooperation Agency, or AMEXCID) led a project to provide the Caribbean with a full and connected geographic information system, including training and infrastructure. Through our participation in UNGGIM, INEGI received an Australian-led capacity building workshop which started and enabled our national geospatial Data Cube project. And ESRI, the leading satellite and geospatial technology company worldwide, has an ongoing partnership with INEGI to provide the Institute (and other institutions in Mexico) with consistent satellite images which are used to build our official cartography.

(b) Updating urban and rural cartography, as well as the National Geostatistical Framework, using satellite imagery and other private data inputs. Within the framework of the geographical activities carried out by the INEGI, and through the international alliances and participations described above, there is an urban and rural cartographic update of the country, which is used for georeferencing censuses, surveys and administrative records. For this, it is necessary to have satellite images of the highest possible spatial resolution, although currently INEGI does not have private collaborators that grant specific rights for its access and use through an agreement, there are services contracted to DigitalGlobe, belonging to the Maxar Technologies group, who provide the high-resolution image service. The mechanism for access to high-resolution images is through Agreements with the Secretary of Navy, which has the EVISMAR acquisition contract with the image provider company, called MAXAR. According to the institutional experience, it has been seen that the best way for successful collaboration with private data providers, would be for government in general to coordinate to acquire high-resolution satellite images in a "consolidated" manner, in this way they will achieve better purchasing conditions (price and, above all, the distribution rights among the participants), in addition this kind of agreement will avoid the duplication of the purchase of the same information.

(c) INEGI currently leads the Business Natural Capital Accounting Working Group (BAWG) within the UN Expert Group on Environmental Economic Accounting, which aims to build bridges between private and public initiatives to share data and make it comparable, as well as develop common terminologies and methodologies. In terms of cooperation with private partners, our participation in these UN groups, led to the development of a pilot project on Natural Capital Accounting and Valuation of Ecosystem Services (NCAVES), coordinated with the UN Statistics Division and the UN Environment Programme, and financed by the European Union.

D. Collaboration with civil society organizations on the value of volunteer work

20. An example of collaboration with the private sector and organized civil society has been the work done within the framework of a technical working group with the participation of the Mexican Center for Philanthropy (CEMEFI), the Center for Research and Studies on Civil Society (CIESC), among others, which periodically conducts information surveys through directories, reports and specialized surveys, with the aim of learning information needs from this sector that official statistics currently do not address. A first approach within this working group made possible to assess and analyze the information collected, which allowed to adjust some statistical programmes, particularly the toolkit of the satellite account of Non-Profit Institutions (CSISFL), that considers the reference framework of the International Labor Organization (ILO) for the measurement of volunteer work, as well as the classification and concepts of the System of National Accounts.

21. Incentives to invite private companies to share private data are diverse and depend on the project. Sometimes, seeing their data as part of official data published by INEGI may be enough incentive for private partners to agree to collaborate. Another important incentive is ensuring the participation of more than one private entity, so that the resulting, compound, information can provide them with greater and better knowledge of their economic sector. Our experience with civil organizations shows another type of incentive to share private data in their possession: adopt and adapt their data to generate information with greater added value through a methodological treatment with which INEGI has extensive experience, for example, quantifications of volunteering work. This can apply to any sort of company or organization, as it does to other public actors: partnering with an NSO provides them with the opportunity to develop methodologies, surveys or indicators with the quality and comparability standards of official statistics, that are adapted to meet their information requirements, and which can be shared and analyzed by all actors.

22. Although economic incentives are an additional possibility, the legal framework does not regulate them, and except for the purchase of satellite images for projects that involve access to private data, there are usually no monetary considerations. However, this is the reason that has made it difficult to carry out other projects, such as mobility studies based on cellphone data.

23. When the project involves the publication of data, the confidentiality of the information of the institution holding the data and of the units (individuals or companies) that make up the data must be ensured, as well as the parity of the conditions and requirements made of all partners within a specific project – such as those involving the banking and automotive sectors.

24. The projects carried out have required resources and the development of capabilities. Until now, the technological needs of the projects have been focused on two aspects: 1) innovation and implementation of secure and flexible mechanisms to transfer and exchange of data and 2) storage and processing capabilities of large volumes of data. Regarding the channels for data transfer, data holders expect to have simple and secure mechanisms, in accordance with their internal policies. In this sense, the adaptation and not the imposition by the statistical offices is relevant. For example, some companies agreed to the installation of software developed by INEGI for file exchange, while others, their institutional policy only allows sending by email. INEGI is currently carrying out a project that will serve to standardize different data transfer mechanisms from external sources. The sole request for

data raises a prior need: to carry it out properly, it is necessary to know the data, its scope and limitations, as well as its generation process. Training in new sources of information and new technologies are necessary from the beginning of the project.

III. Key takeaways and challenges

25. NSOs are likely to face all kind of challenges when approaching private sector representatives to collect data. In this regard, we recommend highlighting benefits for companies and non-government actors when cooperating with NSOs, which can include processing of raw data, timely publication of market outlooks or higher levels of data granularity publicly available in specific industries. This was INEGI's experience when cooperating with the banking sector and the Mexican automotive industry.

26. We have also explored possibilities, comparative advantages, and expertise that we, as an NSO can offer to innovate business models, help companies with data quality, confidentiality, and cybersecurity issues, and develop capacities reciprocally, so that we may build mutual trust and address mutual concerns about data sharing. Partnering with an NSO can provide private actors with the opportunity to develop methodologies, surveys or indicators with the quality and comparability standards of official statistics, that are adapted to meet their information requirements, and which can be shared and analyzed by all.

27. Successfully negotiating with private owners involves knowing the needs of the industry in general and of each of the companies, since not all of them require the same breakdown of information for their decision making. For this purpose, continuous and careful dialogue is important to identify and learn about the information needs and interests of private partners, as well as the timing, frequency and data disaggregation they require, in order to provide useful information for their priorities and those of the NSO.

28. The COVID-19 experience left us many important lessons at the national and international level. For NSOs, it meant exploring and learning new, alternative data sources and adapting our traditional, long processes to make them timelier and more useful to novel or urgent needs – and still ensure quality, comparability and robustness for decision making. In addition, new technologies and actions such as home-office, online shopping, lodging and daily displacement (among other examples), have changed the way of life and the behavior of people and their habits – including rethinking the role and relevance of NSOs. Also, further work needs to be done to create synergies between traditional and non-traditional data sources, including from private sources and administrative records.

29. These experiences, including those described in section II, should encourage NSOs, the private sector and non-government organizations to work together to exchange information and management of data. Partnering up can grant access to new data, highly specialized information, such as data collected by the banking and telecommunications sectors, or specialized geospatial and/or Earth observations information.

30. If done right, these data-sharing partnerships can be win-win interactions – for users, for the economy and to provide better data. Both official statistical information and experimental statistics are important for the generation of more timely information and address new issues that society demands. This is particularly important in cases and circumstances of added public/collective interest, such as during emergencies, for public health reasons and for the monitoring and protection of ecosystems.

31. In Mexico, the legal framework regulating the private sector can be very detailed and complex, as is the case of telecommunications or banking. As a result of our experiences with the private sector, INEGI has carried out an analysis of these legal frameworks and of their main actors, either as regulators or because they are preponderant companies in the market. Through these analyses, we have identified limitations in access to data (for example, due to the impossibility for banks to transfer data with identifiers that may violate banking secrecy laws), as well as the scope of the sector regulators so that, through them, or directly from them, information can be requested.

32. More work needs to be done on legal initiatives to expand the capabilities of NSOs to facilitate the access to new sources of information, which will reinforce the use of official

statistical information and promote stronger strategic alliances with different sectors of society. These frameworks, which result in an effective data governance, both at the national and international level, should be based on clear and strict guidelines and principles, which guarantee the rights interests and benefits of all involved.

33. Despite the many individual examples and success stories that can be found through bilateral, multilateral or other multi-stakeholder alliances and partnerships, not many mechanisms exist for consistent, well-regulated and mutually beneficial sharing of private data, and use of both public and private data. A global data governance, backed by trusted entities and where all interests and rights are guaranteed – including intellectual and commercial property, confidentiality and privacy, and data access for the public good. International organizations, such as the UN and the OECD, and even the European Union, should do more to develop platforms and frameworks for such exchanges – particularly involving multinational companies with presence in (and data from) many countries.
