United Nations ECE/CES/2019/31

Economic and Social Council

Distr.: General
2 May 2019

English only

Economic Commission for Europe

Conference of European Statisticians

67th plenary session
Paris, 26-28 June 2019
Item 2 (a) of the provisional agenda
New data sources – accessibility and use
Session 1: Accessing new data sources

Five experiences in using new data sources in Mexico

Note by the National Institute of Statistics and Geography

Summary

The National Statistical and Geographical Office of Mexico (INEGI), has started to explore the use of nontraditional data sources for producing statistical and geographical information. As many other national statistical offices in the world, INEGI is looking for alternative sources of information, given the problems related to producing data from censuses and surveys. This document briefly discusses five of those experiences. The first one, *The mood of twitterers in Mexico* generates statistical information from georeferenced tweets that are broadcasted daily from any location within the Mexican territory. It reports the "positivity quotient" that relates the tweets with positive emotional load to those that have negative emotional load. The second experience concerns using information from Point of Sale Terminals (POS) and mobile phones to gauge the effects of the September 19/2017 earthquake in México City, both in terms of its immediate consequences for economic activity and on the flows of people out from the areas that were hit most hard by the disaster. The third experience is about using information on electricity consumption by economic units to generate more timely estimations of manufacturing production, based on econometric models that relate electricity consumption with productive activity. The fourth focuses on using satellite images to estimate agricultural production, and environmental consequences of human activity on land cover. This is based on using a time series of satellite images that allows to monitor these phenomena as they occur. Finally, the fifth experience concerns reporting by INEGI of the administrative registers of the Mexican Association of the Automotive Industry with enterprise level data on sales, production and exports.

This document is presented to the 2019 Conference of European Statisticians seminar on "New data sources – accessibility and use", session 1 "Accessing new data sources" for discussion.

GE.19-07176(E)
* 1 9 0 7 1 7 6 *

I. Introduction

1. National Statistical Offices (NSOs) are permanently under pressure to generate more and better information, in an environment of increasing budgetary constraints. In consequence, NSOs are obliged to look for ways to incorporate non-traditional data sources to enrich and improve the efficiency, relevance, timeliness and general quality of their supplied information.

2. Official statistics can be seen as a building supported by 3 pillars: censuses, surveys, and administrative registers. The "data deluge", a by-product of the digital revolution, has generated the possibility of counting with a fourth pillar: the so-called "big data". We still do not know how important big data will become for the production of official statistics, relative to the other pillars, but now we can see that it is, at least, very promising. Even its more sceptical critics now accept that big data should play a role in the production of official statistics. Big data is frequently associated with the three "V": Volume, Variety, and Velocity. Two additional "V"s have been added more recently: Veracity and Value. All these "V"s constitute valid elements for the characterization of a slippery concept, but we think they tend to put most of the weight on data itself, rather than on the way it is used. Even if big data emerged from a series of technological innovations, its essence is more on the side of the way data is approached. Therefore, we can even think about a big data paradigm, according to which big data can be seen as a flexible approach to use and re-use the totality of a data set, structured or not, in a diversity of possible purposes, normally different from those for which the information originated.

II. Description of content

3. Considering the arguments presented above, INEGI has started to approach data from this new paradigm and to explore the use of non-traditional data sources for the generation of statistical and geographical information. This paper briefly reports five of those experiences.

4. The first one, which is called *The mood of the twitterers in Mexico* generates statistical information from georeferenced tweets that are broadcasted daily from any point of the Mexican territory and reports the "positivity quotient" that relates the tweets with positive emotional load to those that have negative emotional load.

5. The second experience concerns using information from Point of Sale Terminals (POS) and mobile phones to gauge the effects of the September 19/2017 earthquake in México City, both in terms of its immediate consequences for economic activity and on the flows of people out from the areas that were hit most hard by the disaster. The POS analysis shows the changes in sales attributable to the natural phenomenon and the time it took to return to the original trend. The mobile phone information allows seeing changes in the patterns of mobility of people attributable to the earthquake, which also enables us to estimate the proportion of people forced by the circumstances to go somewhere else to spend the night.

6. The third experience is using information on electricity consumption of economic units to generate more timely estimates of manufacturing production, based on econometric models that relate electricity consumption with productive activity of the economic units in the sample of INEGI. This information is obtained from the Federal Commission of Electricity (CFE), which is the main electric company in charge of supplying electricity, as well as measuring the consumption of electricity of most of the country. Since CFE measures consumption on a high frequency basis, through the collaboration with CFE, INEGI is able to update its series of manufacturing production and the use of electricity consumption in a timely fashion.

7. The fourth experience focuses on using satellite images to estimate surface statistics related to agricultural production, or environmental consequences of human activity on land cover. This is based on using time series of satellite images, integrated in a spatial data cube, which allows monitoring these phenomena as they occur over space. The use of Machine Learning techniques has permitted to run supervised classifications using INEGI's in-situ data.

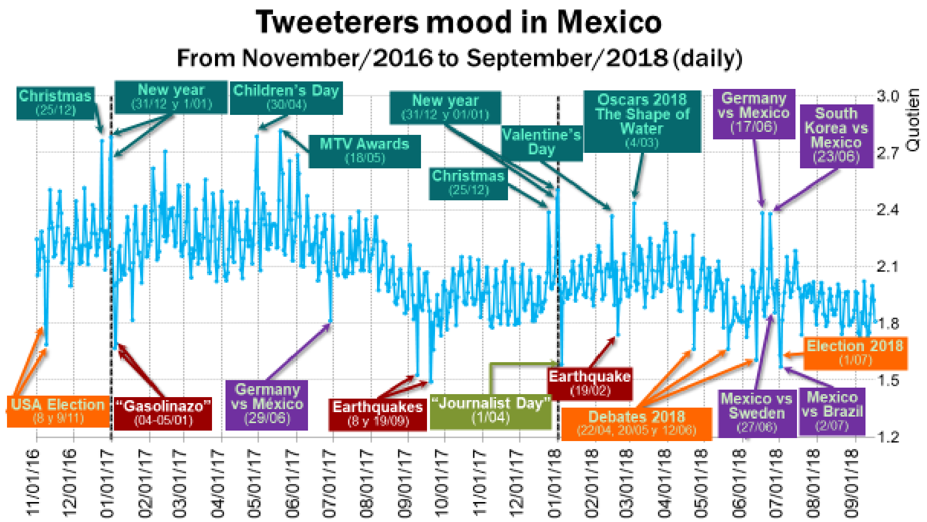
8. Finally, the fifth experience concerns reporting by INEGI of the administrative registers of the Mexican Association of the Automotive Industry, with enterprise level data on sales, production and exports.

9. The following paragraphs describe each of the cases listed above.

A. The mood of twitterers in Mexico.

10. *The mood of the twitterers in Mexico* generates statistical information from georeferenced tweets that are posted daily from any point of the Mexican territory. It reports the “positivity ratio” that relates the tweets with a positive emotional load to those that have a negative emotional load. The process uses sentiment analysis techniques that involve tagging a subset of tweets by human beings and the subsequent machine learning process that allows automating the identification of the underlying emotional charge, which happens to match the human criterion more than 80% of the time. This enables us to generate the highest frequency time series published by INEGI, since every day the computer updates the classification of about 100,000 tweets with information up to the previous day. The system, freely available on Internet, allows consulting at national and regional level in annual, quarterly, monthly, weekly, daily and even hourly series. Also, it is possible to see the hashtags that dominated the scene and directly access news sites that help to associate the variations in the series with specific events. With a time series that begins in 2016, the system shows dynamics consistent with what can be expected for a number of clearly identified events. Such events are, for example, the results of the US Presidential Elections on November 8th, 2016 (negative); the increase in national petroleum prices at the beginning of 2017 (negative), and the two large earthquakes of September 2017 (both negative). Furthermore, it provided valuable insights about the triumph in Soccer of Mexico against Germany on June 2017 (positive), and the Oscar award for "The Form of Water" as best film of 2018 (positive). This is a practical exercise in the generation of statistical information on inferred aggregate subjective well-being using entirely non-conventional statistical information that can be refined to address particular topics in relation to how they are felt by the twitterers. This product is the result of combined research efforts of INEGI and national and international academic institutions and has been the basis for developing other data science projects at INEGI.

Figure 1
Twitterers mood in Mexico



11. To build the *Mood of the Twitterers* it was necessary to download hundreds of millions of tweets georeferenced to the Mexican territory, available through the free Twitter API[1]. We started by manually labelling a relatively small set of training tweets. Human labelers

[1] Application Programming Interface.

(taggers) were asked to gauge the underlying emotional charge (positive, negative or neutral) of each tweet. The resulting set of tagged tweets was then used to train a computer to replicate the human criterion needed to classify tweets in two categories: positive and negative. The training set was composed of about 54,000 standardized tweets, tagged by 5000 students of University TecMilenio and 5000 by our INEGI co-workers. Both sets of taggers were scattered all over the 32 federate entities (states) of the country. Human taggers from each state received tweets generated in the same state in such a way as to facilitate the proper interpretation of regionalisms. Each tweet was assessed by different taggers and each tagger evaluated the same tweet several times, so that information from inconsistent qualifiers could be identified and disposed of. Additionally, the underlying emotional charge of each tweet could be identified as each tweet was systematically qualified by different people as positive, negative or neutral, which helped to achieve a more robust qualification.

12. The tweets classified by humans were then used to train the computer, for which it was necessary to incorporate an assembly of classifiers developed with the support of experts in data science from INFOTEC and Centro Geo, two outstanding Mexican research centers. After comparing many alternatives, it was found that an assembly of 33 different Support Vector Machine (SVM) estimation procedures, each one run with a different 80% of the tweets tagged by humans and available for training and validation[2].

13. Using a set of tweets classified by humans that were not used in the training of the machine, it was possible to establish that such assembly of classifiers allows the computer to properly classify 80 of each 100 tweets, which is a percentage of particularly high success among experiences in the field of sentiment analysis. Once with the computer properly trained to classify tweets, INEGI was given to the task of exploiting the results and presenting them. It should be noted that although the information published on Twitter is public, INEGI only reports aggregate data and at no time reports nominative or individualized tweets. Even in the tweets classification system (Pioanálisis) the tweets were anonymized before they were presented to the human taggers. So, every day INEGI reports the tweets from the day before, automatically classified by the computer. To do this, the computer simply uses the 31 SVM hyperplanes previously estimated and decides the positivity or negativity of each tweet according to a majority report rule.

B. Point of Sale Terminals (POS) and mobile phones to gauge the effects of an earthquake in Mexico City

14. On September 19th, 2017, an earthquake of magnitude 7.1 on the Richter scale and epicenter in the State of Morelos (just 120 kilometers from Mexico City) struck Mexico, causing human and material losses in various states of the country.

15. According to information released by public institutions and the media, the total number of fatalities was 369, Mexico City being the entity with the most victims. Thousands of homes were partially or totally damaged, forcing its inhabitants to change their place of staying.

16. In an inter-institutional collaboration effort, BBVA Bancomer, Telefónica Movistar, UN Global Pulse and INEGI agreed to estimate the economic and mobility effects of the earthquake, both in Mexico City and in the Roma-Condesa corridor, one of the most affected areas inside Mexico City. This established the basis and mechanisms for developing reaction protocols in matters of public and private information that are integral, pertinent and timely. There are several examples of protocols that have been promoted internationally by UN and

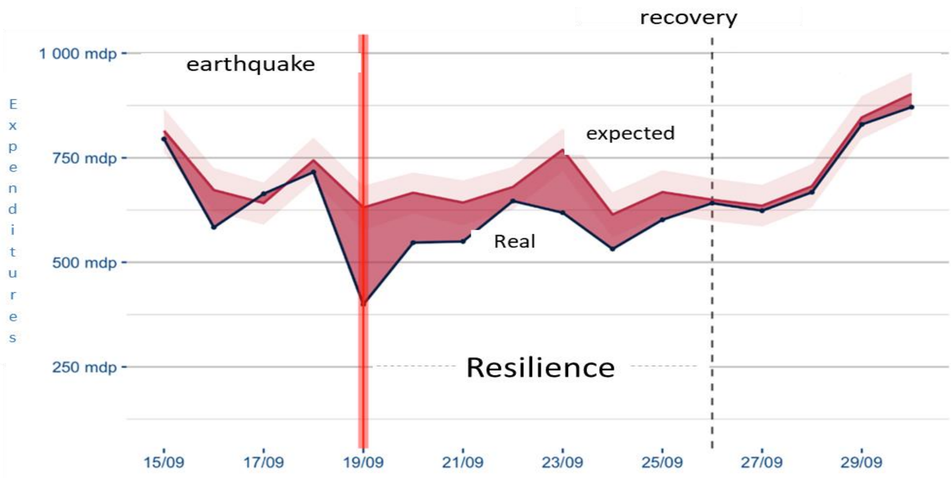
[2] The classification was not run over words but instead, we used q-grams of different lengths. In fact, we found optimal to use q-grams of order 3,4,5 and 7. We also normalized text, gave polarity values to emoticons, implemented number, URL and user substitution, and transformed words to basic linguistic roots. TF IDF scheme (TFIDF, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus) was used to achieve vectorization necessary for the classification. This allowed us to identify clusters of tweets based on chains of q-grams of different orders derived from each tweet. Since we trained the computer with a set of tweets tagged by humans, we were able to use SVM to estimate the multidimensional hyperplanes that best separate tweets that we know were positive from those that we knew were negative.

derived from global projects, such as the Sustainable Development Goals, the Sendai Framework for Disaster Risk Reduction 2015-2030 and more specialized documents such as the Technical Guidance for Monitoring and Reporting on Progress in Achieving the Global Targets of the Sendai Framework for Disaster Risk Reduction.

1. Economic impact

17. We used a methodology that takes as inputs the data registered in point of sale (PoS) terminals within Mexico City, as well as data from states not affected by the earthquakes. Based on this information, we generated as a counterfactual a predictive model for Mexico City using the commercial activity from PoS terminals of states not affected by the earthquake.

Figure 2
Economic impact



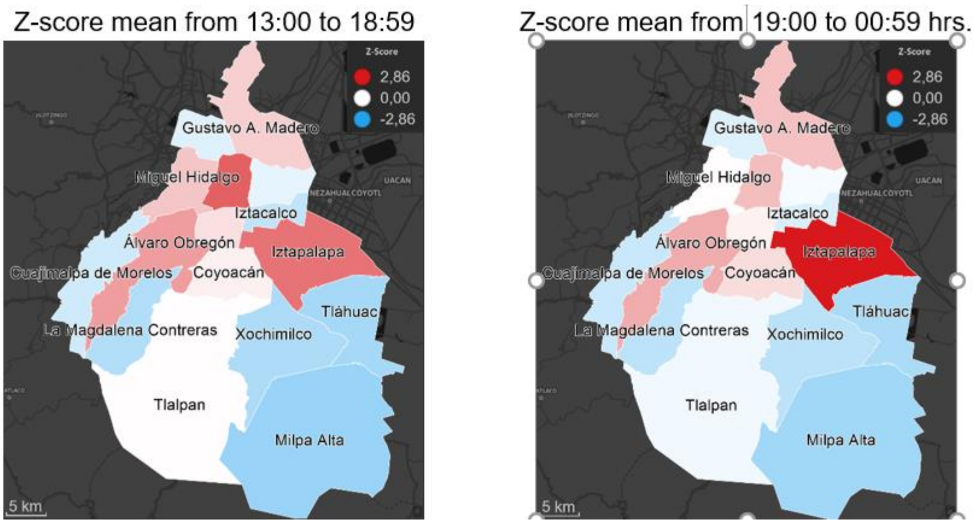
18. If we define the economic impact as the difference between the estimate of the model (counterfactual) and the observed data, the fall of September 19 signified a decrease of 36.7% in the registered expense with respect to what was expected. For the entire resilience period (7 days), the decrease was 14.7%.

19. The effects on both the number of victims and physical damage were concentrated in certain geographical areas of Mexico City. One of the most affected was the Roma-Condesa corridor. Within the zone, the estimated recovery time was 6 days and the losses meant a 21.8% fall in the expected expenses. It is worth mentioning that for some economic activities, such as Entertainment or Restaurant & Bars, the resilience took longer (more than 20 days).

2. Impact on mobility

20. We use the CDR data (Call Detail Record) of Telefónica Movistar to approximate the mobility effects caused by the earthquake of September 19th. To identify unusual patterns from CDR data, a standardized empirical distribution of displacements is generated for each day and hour of dates prior to the earthquake; that way the displacements for each of the hours after the earthquake can be compared with the usual pattern. Thus, Figure 2 shows in more intense colors (either red or blue) the ‘Alcaldías’ (municipalities) with the greatest deviations from the number of regular trips made from that Alcaldía to any other. The map on the left side shows the behavior during the 6 hours after, that is, from 1:00 p.m. to 6:59 p.m. while the map on the right side shows the behavior from 7:00 p.m. to 00:59 p.m. the next day.

Figure 3
Deviation from the common pattern of trips



21. It is interesting to observe in the previous figure the differentiated impacts on mobility: in the first six hours after the earthquake, the Alcaldía Miguel Hidalgo (one of the most affected) recorded the biggest impacts, meaning people displacing more than usually from that Alcaldía to any other. The Alcaldía Iztapalapa was another with important effects on mobility as it was expected much more displacement than what actually occurred. Lastly, there were Alcaldías like Tlalpan behaving as expected.

C. Estimation of the manufacturing economic activity index based on electric power consumption data

22. The purpose of this work is to make use of relevant administrative records from the Federal Commission of Electricity (CFE), associating it with INEGI's business directory (*Registro Estadístico de Negocios en México* – RENEM in Spanish), to generate an electric power consumption index (ICEE). The main advantage of ICEE is that, since it is updated 15 to 20 days after the end of the month and presents a strong linear correlation with the manufacturing sector industrial activity index (IMAI 31-33), it can be used to accurately anticipate the IMAI´s 31-33 current month observation, before the official data is released.

23. The process to generate ICEE, and to validate its relationship with IMAI 31-33, is the following:

(a) Preprocessing CFE data: an imputation analysis is performed. CFE data is complemented with information form the National Energy Control Center; outlier identification is carried out and data is matched to be incorporated into the index construction.

(b) ICEE construction: for each month, weighted electric power consumption data for each registry is associated between CFE-RENEM. Weighting is made upon the manufacturing subsector. The weights are provided by Mexico's National Accounts System (SCNM), in order to get an appropriate weighted estimation, which is then re escalated to make it comparable to IMAI 31-33.

(c) Regression model: a regression model is estimated with first order auto-correlated errors through the Cochrane-Orcutt method, where IMAI 31-33 is the dependent variable, and ICEE is the independent one. Additional dichotomous variables are also incorporated. Given the nature of the regression, the interest variables are included in first differences.

(d) Model validation: Durbin-Watson statistics are calculated to verify first order auto-correlation absence; Breush-Pagan test is performed to corroborate lack of heteroscedasticity, and finally, residuals' normality is validated through Shaphiro-Wilk test.

(e) Nowcasting: early updating the information and using the regression model, we estimate the IMAI´s 31-33 current month observation, both in levels and in annual and monthly differences. The results indicate that the mean of the absolute error is 1.53 units in levels, and 1.23 for the series in monthly variations during 36 months (January 2016 to December 2018).

24. Future lines of investigation consider the incorporation of additional information, smooth the series via seasonal adjustment, look for relationships among different IMAI 31-33 disaggregation levels, as well as ICEE construction for different levels of economic activity.

Statistical validation

25. The analysis allows for the timely estimation of IMAI 31-33. Note that, the first differences regression with first order auto-correlated errors, is similar to conduct a cointegration exercise, since the one period lagged residual is incorporated into the differentiated variables. From this perspective, the assumption of potentiating the use of cointegrated time series is fulfilled. Additionally, the work sticks to SCNM's normativity and the assumptions of the statistical modelling are validated by the data. In conclusion, the timely information that the proposed ICEE allows, has helped INEGI to precisely anticipate IMAI 31-33.

D. Mexico's Geospatial Data Cube

1. Satellite data as input to INEGI's geographical and statistical products

26. At a general level, INEGI has produced maps of Mexico for several decades. Initially in paper format (1970s, 1980s), and now in the form of themed geospatial datasets. In previous years, aerial photographs were used as input to map production, but today, the use of satellite data in particular has become the primary input for natural resources maps. However, the current process of extracting information from satellite imagery is operator-assisted, and labor intensive. This includes: inspection, analysis, feature extraction (as in identifying roads, rivers or water bodies), and interpretation (which involves deeper knowledge about the relationships among variables and processes involved).

27. Satellite images are not used as input in the Agriculture Statistics sector of INEGI, but they do ease fieldwork carried out in censuses and surveys related to it. However, research has been done for identifying the coverage of specific crops using satellite images. Results of these practices are: (i) the construction of a field verified database and (ii) its application for classifying bigger areas; the latter enables the experimental statistical estimate of area per crop type.

28. Besides the above-mentioned, data traditionally used for Geospatial Information production, like field data, often suffers from inconsistencies (changes in reporting methodologies or missing measurements), EO[3] Satellite data provides a unique opportunity to build consistent time series and hence, implies better monitoring.

29. However, regarding the use of satellite images as source data, two drawbacks found in the current methodologies and infrastructure are: (i) insufficient resolution of satellite imagery for specified functions, and (ii) the labor and time required to extract information from the data and to integrate it with the geographic register, or products. Therefore, as the demand for more frequent and detailed map products increases, issues such as data storage, life-cycle management, and processing large and varied volumes of satellite images need to be addressed.

2. Mexico's Geospatial Data Cube enabling efficient EO use

30. The latest machine learning techniques help to overcome challenges like labor and time demanding processes by properly training the software with the guidance of human experts. Combined with the right Big Data infrastructure, insufficient spatial resolution of

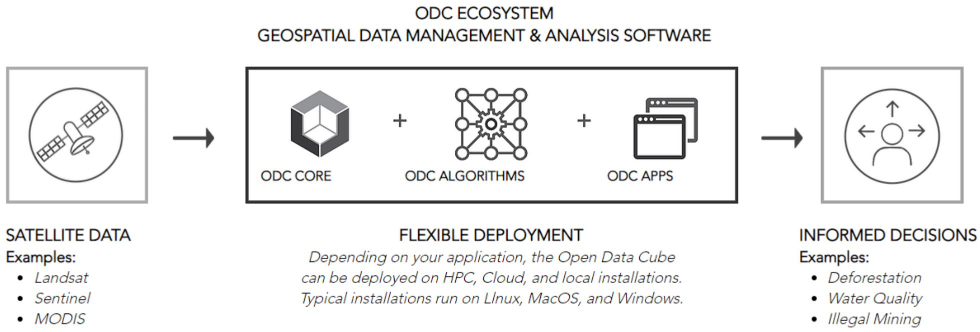
[3] Earth Observation.

satellite imagery might be addressed by increasing temporal resolution: using multiple images in the analysis, instead of limiting the data source to a single one. These approaches are all found in a geospatial data cube, which is a massive array of multidimensional raster data that allows for smart and efficient handling and analysis of the immense volumes that constitute satellite data.

31. Through a collaboration with Geoscience Australia, the Mexican Geospatial Data Cube was developed at INEGI. This tool (Open Data Cube) has enabled exploiting Earth Observation data in a timelier manner worldwide.

32. The Mexican Geospatial Data Cube features a nationwide database for archive and recent satellite images that are previously standardized to allow processing information regardless of the source sensor. It also spatially aligns the images; this enables digital time-series analysis of this database at pixel level, instead of relying solely on visual inspection. Furthermore, since the minimum analysis unit is now a pixel, high-cloudiness images are included in the database, which means the remaining non-cloud pixels can be used (most processes used to dismiss such images).

Figure 4
ODC Ecosystem Geospatial Data Management & Analysis Software



Source: OpenDataCube.org

33. In summary, developing this more coherent approach will propel the processes involved in exploiting satellite images in INEGI: data can now be stored, managed, accessed, and analysed more efficiently.

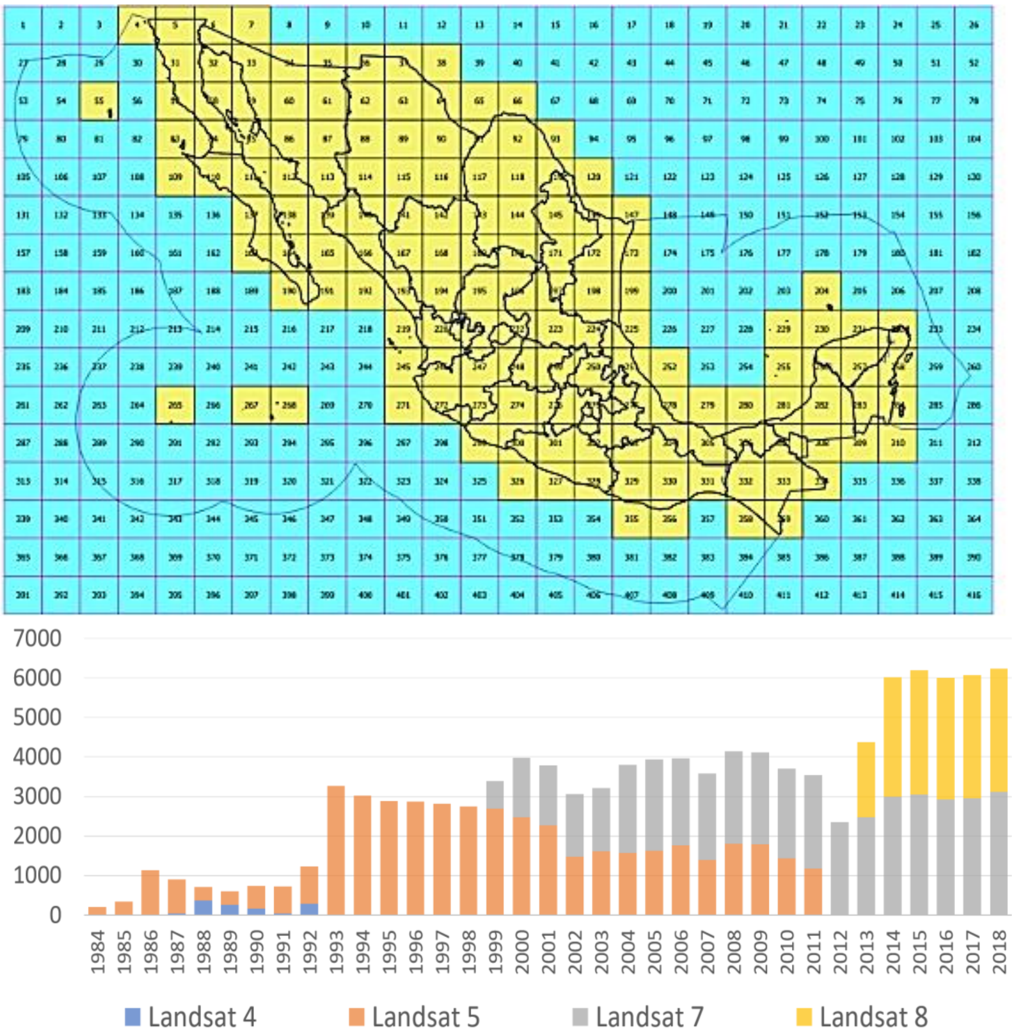
3. Institutional level implementation progress (and paradigm shifts)

34. INEGI has already received the complete ARD[4] Landsat archive with over 109 thousand images, which was processed and delivered by USGS/NASA staff; it has a volume of 30TB. This archive is loaded into the institutional infrastructure of INEGI, and is estimated to increase its volume to 90TB, when decompressed.

[4] Analysis Ready Data.

Figure 5

Landsat ARD Archive (top) & Title grid designed by INEGI (bottom)



35. Another step towards the implementation of Mexico's Geospatial Data Cube was defining a suited grid and projection for the Landsat ARD archive; this grid grants data compatibility for other data source resolutions, like Sentinel. Likewise, INEGI favoured Albers equal-area projection; as implied by the name, it grants projecting all areas proportionally. This will provide easier surface related statistics; which is a common approach for geospatial-based SDG indicators. This grid also ensures covering Mexico´s exclusive economic zone.

36. Additionally, INEGI is designing the services to bring end users closer to EO data analysis, which further fosters the open-data philosophy: both images and algorithms will be made available to the user.

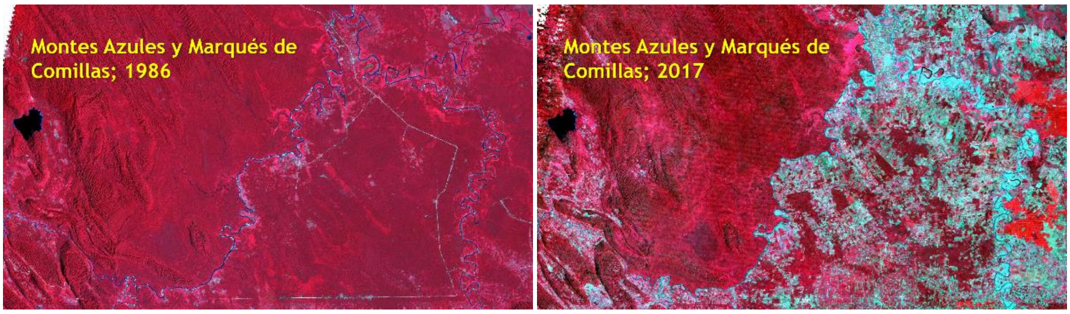
4. Application examples

A. Vegetation evolution

37. Although quite simple, this example shows the importance of monitoring natural resources throughout archive files, so that we can better understand our problems and current situation. Vegetation evolution was analyzed, in time in the area of Montes Azules and Marqués de Comillas, comparing Landsat observations from 1986, and 2017. During these 3 decades, deforestation has had quite a bigger presence to the right of the river. This result may regard assessment of public policy's effects too, since the Usumacinta river acts as the limit for the natural protected area of Montes Azules. The regions compared can be appreciated in the images below.

Figure 6

Comparative images of Montes Azules de Comillas in 1986 and 2017



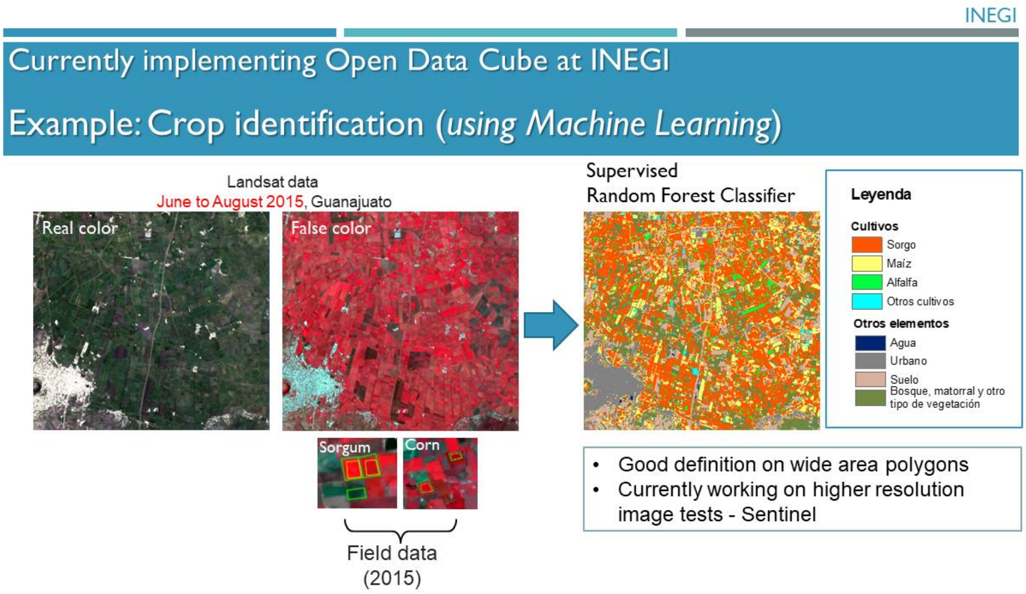
Source: INEGI, at the GEO-XV Plenary Session 2: EO in Support of the SDG (Kyoto, Japan, October, 2018) Vegetation evolution example in the area of Montes Azules and Marqués de Comillas using 1986 and 2017 Landsat data *(Rendering: false colour or pseudo colour)*
Source: INEGI, at the UNWGIC Parallel Session: Measuring and Monitoring the SDGs (Deqing, Zhejiang Province, China, November 2018)
INEGI intends to technically describe the procedure, computational sizing and tools to apply the UN methodology (IPCC) for indicator 15.4.2 using ODC

B. *Crop identification*

38. Another example for the use of the Data Cube in INEGI is crop identification, using machine learning techniques. A supervised classifier was applied using INEGI's field data from 2015 and Landsat data from June to August of the same year. Results are satisfactory in the polygons of bigger areas. Due to the presented confusion in the results for smaller polygons, tests are currently being recreated using higher spatial and temporal resolution images (Sentinel-2). Further work and planning of research initiatives in INEGI's Agriculture Statistics area, met the conclusions that the Open Data Cube and the use of Sentinel-2 data offer a variety of advantages in agriculture related products, which are summarized in the annex table.

Figure 7

Crop identification example using 2015 Landsat images and field data (with Random Forest Algorithm from the Open Data Cube)



Source: INEGI, at the GEO-XV Plenary Session 2: EO in Support of the SDG (Kyoto, Japan, October 2018)

E. Data on automobile industry

39. As already said, the complexity of statistical development needs more sources, either from census/surveys, from administrative records or from big data. The administrative data play an important role in completing the statistical information, while helping to reduce the respondent's burden.

40. Unlike the compilation and treatment of the administrative records of governmental entities, INEGI recently obtained the authorization of the 21 companies affiliated with the Mexican Association of the Automotive Industry (AMIA) and Mitsubishi Motors to publish information on the number of vehicles of each of the 33 brands and models that the corresponding companies produce and / or market in Mexico.

41. It is important to stress that this information provided directly by the companies, is not subject to any statistical treatment and maintains the nature of administrative registration. Therefore, when a company does not provide the required information, INEGI uses an N.D. (Not Available) tag, indicating that the figure was not delivered to INEGI for publication.

42. Despite the complexities of this administrative data, INEGI is taking the first steps in a broader direction that in the following years will be improved.

43. Nowadays, the Administrative Registry of the Light Vehicle Automotive Industry offers monthly data by company regarding the sale, production and export of new vehicle units, without specifying monetary values.

44. As of October 2018, INEGI initiated the disclosure of the Administrative Registry of the Automotive Industry, maintaining the public character of the information that, from the eighties until September 2018, the AMIA had disseminated. In the past, AMIA used to concentrate all the participants in the market. So, the participation of INEGI in the reporting of the automotive industry allows for maintaining the completeness of the reports.

45. The incorporated theme and the way in which this administrative record is presented coincide with the criteria followed by AMIA that have been previously agreed with the companies.

46. INEGI pursues the goal of maintaining and publishing the administrative registry of the light vehicle automotive industry, corresponding to the sales, production and exports of automobiles and light trucks in Mexico. It also tries to present the information about the vehicle units at the brand and model level, identifying the segment and country of origin or destination, to contribute with users to the knowledge of the automotive industry.

47. INEGI disseminates the subject matter of this administrative record twice a month, as follows: Total sales to the national market by brand, two to three business days after the reference month. The information is broken down into sales on the domestic market, production and exports broken down by brand, model, characteristics of the vehicle units and country of origin or destination, as well as electric and hybrid vehicles, four to five business days after the reference month.

48. The aspects described above present national information of the immediately previous month, except the commercialization of hybrid and electric vehicles that present a breakdown by federative entity, whose data have a lag of two months with respect to the reference period.

49. It is necessary to point out that on each of the days scheduled for the publication of this administrative record, updates will be made based on the lagged or revised information provided by the companies. The specific dates for the publication of the Administrative Register of the Light Vehicle Automotive Industry are established in the schedule that has been defined in agreement with the companies.

50. The following table shows the harmonization of the last fourteen years:

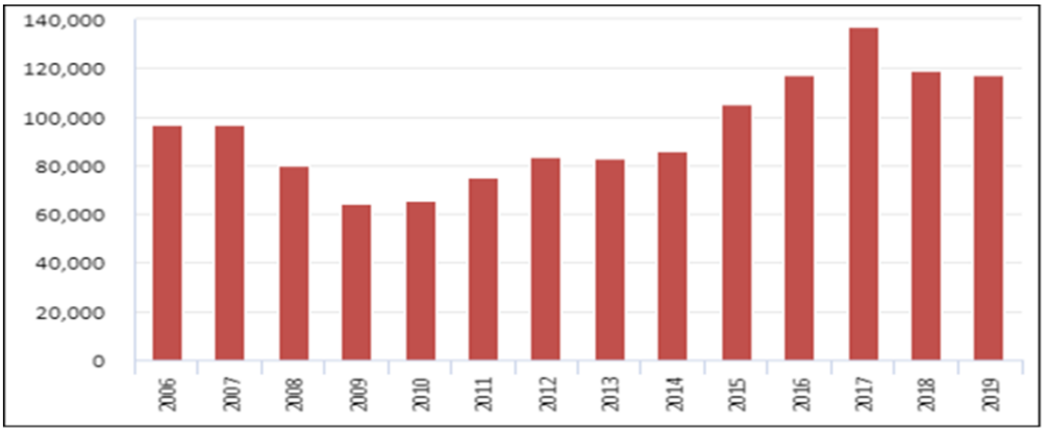
Figure 8

Total sales of light vehicles (including imports)

AÑOS	JAN	FEB	MARH	APR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DEC	TOTAL
2006	96,227	89,079	96,871	77,879	86,462	87,084	83,069	90,937	92,083	97,469	102,201	140,375	1,139,736
2007	97,675	86,060	96,487	75,020	84,756	80,462	83,105	88,573	86,547	97,182	97,694	126,329	1,099,890
2008	96,846	86,997	80,119	83,106	85,827	81,424	85,324	86,119	76,620	83,307	78,555	101,300	1,025,544
2009	69,664	61,579	64,242	51,395	53,440	55,974	56,443	58,926	58,505	67,882	64,914	91,961	754,925
2010	64,064	59,518	65,414	60,432	61,632	59,910	61,960	66,931	65,934	74,095	75,582	104,941	820,413
2011	68,767	66,990	75,125	65,246	68,634	68,366	68,533	75,681	73,998	75,748	83,107	115,698	905,893
2012	75,297	74,704	83,574	69,890	80,268	78,508	76,378	83,326	79,961	83,172	91,966	110,998	988,042
2013	84,403	80,285	82,860	83,647	87,638	83,858	86,760	88,586	78,555	88,416	100,571	119,519	1,065,098
2014	85,614	80,037	85,767	76,941	88,388	84,207	96,366	103,994	89,313	101,090	111,837	133,411	1,136,965
2015	103,805	97,659	105,034	94,953	102,156	107,097	111,863	112,307	111,705	120,214	126,750	160,901	1,354,444
2016	119,833	111,126	117,252	118,754	121,879	134,913	132,109	134,388	131,888	137,503	154,779	192,741	1,607,165
2017	123,447	118,193	137,245	114,907	123,381	127,695	122,626	125,916	116,654	123,526	141,616	159,110	1,534,316
2018	109,264	109,698	118,907	109,534	114,933	120,069	114,811	119,222	114,653	117,325	133,791	141,963	1,424,170
2019	111,212	103,679	117,122										332,013

Figure 9

Total Sales of Light Vehicles at each March for the years 2006-2016



IV. Conclusion

51. The Mexican national statistics office, as many others in the world, is looking for alternative sources of information given the problems related to generating data from censuses and surveys. These problems range from high costs to no answer, and mean a diminished capacity of the governments to understand the underlying forces that cause economic growth and social development. This document presents five examples of the use of these new sources of information.

52. The advances in technology have facilitated the use of new sources of data to understand economic and social processes in a more timely and efficient manner. This is a new way to tackle old and relevant questions related to, for example, the economic activity of a country and its regions, or the social processes contributing to the wellbeing of the population. However, these new sources of information can also respond to new questions regarding diverse phenomena, like mobility after a natural disaster or the sentiment of the population facing a certain event.

53. INEGI is highly committed to the generation of relevant information for decision making, either at the government level or in private spheres. This represents a challenge in terms of generation and use of data sources. We will continue to pursue excellence through agreements with other entities and by the use of big data. Although this is relatively new for us, the examples presented here have sowed a seed that will definitely pay off in the future.

Annex

Comparing traditional methodologies vs. methodologies using ODC for agriculture related activities

Activities	*Without ODC*	*With ODC*
Crop surface estimation	Download and preprocess images Very likely presence of clouds during the Spring-Summer period (rain season) Likely presence of clouds during the Autumn-Winter period Only one image is used for the estimation Software for processing the image	ARD images Cloud-free images Several images used (time-series) for the estimation Processing algorithms Possibility to develop or customize algorithms
Crop monitoring	Download and preprocess images Very likely presence of clouds during the Spring-Summer period (rain season) Likely presence of clouds during the Autumn-Winter period Only one image is used for the estimation Software for calculating NDVI	ARD images Cloud-free images Change monitoring (time-series) NDVI is featured product, and includes other statistical values (min., max., mean)
Identification of infrastructure for protected agriculture and Verification and update of terrains (parcels)	Download and preprocess images Very likely presence of clouds during the Spring-Summer period (rain season) Likely presence of clouds during the Autumn-Winter period	ARD images Cloud-free images Several images used (time-series) National cover (and nation- wide processing)