

# Cubo de datos **geoespaciales** para el uso de las imágenes satelitales **en la generación de información geográfica y estadística**

## *The Use of Satellite Imagery in the Generation of Geographic and Statistical Information*

**Olivia Jimena Juárez Carrillo, Paloma Merodio Gómez, María del Socorro Ponce Medina, José Luis Ornelas de Anda y Abel Alejandro Coronado Iruegas\***

Los satélites brindan mediciones a intervalos regulares, lo que presenta una oportunidad única para construir series de tiempo consistentes y derivar su análisis en tendencias de comportamiento. Gracias a esta periodicidad, sabemos que el nivel del mar realmente está subiendo, que las selvas tropicales están desapareciendo y con qué rapidez sucede todo esto. Con cientos de imágenes gratuitas liberadas de forma anual, la falta de datos ya no es una limitante; sin embargo, debido a factores como su volumen y complejidad, siguen siendo subutilizadas. En este sentido, este artículo presenta el *Cubo de Datos Geoespaciales de México*, herramienta que busca aprovechar el potencial de dichas imágenes para la generación de información.

**Palabras clave:** cubo de datos geoespaciales; imagen satelital; observaciones de la Tierra; *Landsat*; datos listos para el análisis.

Recibido: 3 de septiembre de 2019.  
Aceptado: 1 de junio de 2020.

Satellites provide measurements at regular intervals, which presents a unique opportunity to build consistent time-series and derive this analysis into behavioral trends. Thanks to this periodicity, we know that the sea level is really rising, or that the rainforests are disappearing and how quickly all of this is happening. With hundreds of free images released annually, the lack of data is no longer a limitation; however, due to factors such as volume and complexity, these images remain underutilized. This article presents the *Mexican Geospatial Data Cube*, a tool that allows analyzing and storing satellite images in a more efficient way, seeking to exploit their potential in the generation of information.

**Key words:** Geospatial Data Cube; satellite image; Earth observations; Landsat; analysis ready data.

\* Instituto Nacional de Estadística y Geografía, jimena.juarez@inegi.org.mx, paloma.merodio@inegi.org.mx, ma.ponce@inegi.org.mx, jose.ornelas@inegi.org.mx y abel.coronado@inegi.org.mx, respectivamente.



Planet Earth against black background/James Cawley/Getty Images

## Introducción

La economía global está viviendo una revolución tecnológica. La explosión de datos es resultado de la traducción de gran parte de nuestras vidas a medios digitales. Las tecnologías que antes eran de uso exclusivo de los gobiernos y los servicios militares se han democratizado; aplicaciones para monitorear tráfico, pronósticos meteorológicos y servicios de transporte son ejemplos del uso cotidiano.

Es probable que esto tenga un efecto relevante en la generación de información geográfica y es-

taadística. Cada día se utilizan más los datos masivos generados en medios digitales para examinar de manera rigurosa la producción de indicadores, como cambios en salarios, productividad, vegetación, etcétera. Con fines de política pública, la información obtenida de dichas fuentes es crucial para la planeación y el seguimiento. Para atender fenómenos naturales, monitorear el crecimiento económico y planear ciudades inteligentes hay, hoy, un sinfín de aplicaciones tecnológicas.

Las oficinas nacionales de estadística se enfrentan a nuevos retos debido al desarrollo de tecnologías que han desatado un *boom* en la generación,

manejo y análisis de datos. Técnicas tradicionales de generación de estadística, como los censos y las encuestas, se enfrentan a los siguientes desafíos:

- Menor disposición de los informantes para proveer datos debido a desconfianza en las instituciones.
- Mayores costos de levantamiento de datos en campo. Por los altos índices de inseguridad, esta labor se ha vuelto muy costosa, comparada con el trabajo de gabinete que permite la era digital, aunada a un menor costo de almacenamiento y procesamiento.
- Los usuarios se han vuelto más exigentes requiriendo información con una mayor temporalidad y, en algunos casos, en tiempo real; por ejemplo, para algunos ya no basta tener información económica anual, buscan tener indicadores que les permitan seguir el pulso de la economía día con día.

Debido a lo anterior, existe una necesidad imperante para desarrollar metodologías que nos permitan generar información de calidad a partir de fuentes no tradicionales, que son aquellas que nos proporcionan datos provenientes de fuentes digitales. Las aplicaciones de los celulares, las facturas electrónicas, los satélites que orbitan alrededor de la Tierra, entre otros, están creando volúmenes de datos cada vez más significativos con cobertura global. La falta de estos ya no es un factor limitante.

Este artículo presenta el *Cubo de Datos Geoespaciales de México*, una herramienta que permite analizar y almacenar las imágenes satelitales de una manera más eficiente, buscando aprovechar su potencial para obtener información.

## **Imágenes satelitales para la generación de información geográfica y estadística**

Este tipo de productos proporciona un punto de vista único para recopilar información esencial para evaluar los peligros ambientales, gestionar los recursos naturales y mejorar nuestra compren-

sión del planeta. Desde que los satélites llevaron cámaras a órbita, se tiene una perspectiva inigualable, que se ha ido enriqueciendo a medida que los sensores avanzan en términos de variedad y sofisticación.

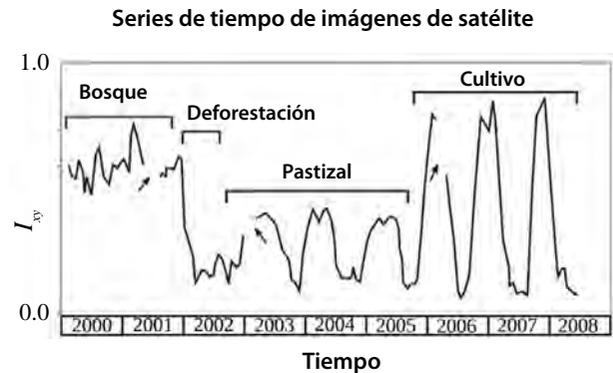
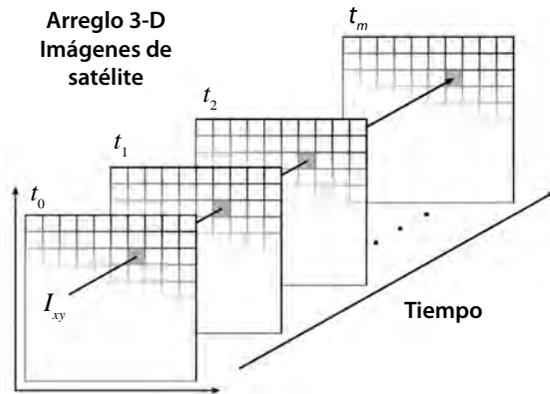
En la década de los 70 del siglo pasado, el Servicio Geológico de los Estados Unidos de América (USGS, por sus siglas en inglés) comenzó a tomar una imagen de la Tierra a través de sus satélites *Landsat* cada dos semanas. Si se reúnen los datos que captaron de México solo durante el 2015 y los almacenáramos en disquetes, se podrían apilar uno sobre otro hasta formar una torre de 1 630 metros de altura, casi la que tiene el Cerro de la Silla en Monterrey, y aprovecharlos para un análisis requeriría de cerca de ocho *laptops* de uso común, equivalentes a 7.5 *terabytes* (TB).

Conforme la tecnología evolucionó, la frecuencia de la generación de las imágenes satelitales también fue aumentando. Con un programa de más reciente lanzamiento (2015), la Agencia Espacial Europea genera una imagen del planeta cada cinco días con los satélites *Sentinel*, lo cual equivale a más de 70 imágenes de la Tierra cada año. Otro ejemplo es la empresa Planet, que cuenta con una constelación de más de 130 nanosatélites y toman una imagen de cada punto de la Tierra al día.

## **Información espectral, mediciones periódicas, datos abiertos**

Los sensores de los satélites juegan, también, un papel en la relevancia de las imágenes satelitales como fuente de información. Así como una fotografía capta la porción del espectro de luz que es *humanamente visible* (rojo, verde y azul), el sensor de un satélite tiene la capacidad de captar, además de la luz, información infrarroja, que es imperceptible para el ojo humano. Uno de los usos comunes de las imágenes satelitales es aprovechar los datos del espectro infrarrojo para conocer información sobre la vegetación, ya que las plantas reflejan la luz en este espectro en proporción mucho mayor que en el verde (ver imagen 1).

## Medición del Índice de Vegetación Mejorado (EVI) a lo largo del tiempo del pixel $x, y$



Fuente: Picoli *et al.*, 2018.

De manera adicional, los satélites brindan estas mediciones a intervalos regulares, por lo que representan una oportunidad única para construir series de tiempo consistentes y derivar este análisis en tendencias de comportamiento. Gracias a su periodicidad, sabemos no solo que el nivel del mar en realidad está subiendo, que las grandes plataformas de hielo se siguen rompiendo o que las selvas tropicales están desapareciendo, sino con qué rapidez están sucediendo.

El gran desarrollo en la tecnología empleada en las misiones espaciales y en los sensores ha revolucionado la industria geoespacial; ahora existen empresas con satélites del tamaño de una caja de zapatos, con sofisticados sensores. Estos cambios y avances retan a generar nuevos productos y mecanismos para el propio análisis de las imágenes.

Sin embargo, la ventaja clave que impulsa el uso a nivel global de las imágenes satelitales provenientes de los programas *Landsat* del USGS y *Sentinel-2* de la Agencia Espacial Europea dentro de las oficinas que generan información oficial es el hecho de que estas imágenes son *datos abiertos*,

lo cual significa que están disponibles en repositorios de acceso libre y gratuito para cualquier agente interesado.

Estos dos programas tienen diferencias fundamentales, como su antigüedad: *Landsat* inició su captación en la década de los 70, mientras que *Sentinel-2* comenzó en fecha mucho más reciente: 2015; otro diferenciador es el volumen de sus datos, a pesar de la posible intuición dada por la diferencia de antigüedad de los programas; gracias a la mayor frecuencia y capacidad de sensores, *Sentinel-2* ha generado un volumen que supera ya el de toda la historia de *Landsat*. Debido a estas características, ambos son fuentes de datos abiertos que se consideran complementarios entre sí.

### Uso actual de las imágenes satelitales en el INEGI para generar información

El Instituto Nacional de Estadística y Geográfica (INEGI) tiene la tarea de monitorear fenómenos ambientales, socioeconómicos y demográficos

que se presentan en todo el país. En particular, el artículo 26 de la *Ley del Sistema Nacional de Información Estadística y Geográfica* indica que: "... El Subsistema Nacional de Información Geográfica y del Medio Ambiente, en su componente geográfico, generará como mínimo los siguientes grupos de datos: marco de referencia geodésico; límites costeros, internacionales, de las entidades federativas, municipales y las demarcaciones territoriales de la Ciudad de México; datos de relieve continental, insular y submarino; datos catastrales, topográficos, de recursos naturales y clima, así como nombres geográficos. A este componente también se le denominará Infraestructura de Datos Espaciales de México...".

Siendo la agencia cartográfica nacional de México, el INEGI ha producido mapas del país durante varias décadas. De inicio, en formato impreso (décadas de los 70 y 80) y ahora, en forma de conjuntos de datos geoespaciales temáticos. Con anterioridad, estos mapas se derivaban de fotografías aéreas, pero en la actualidad, las cartas y datos geoespaciales se obtienen a partir de imágenes satelitales.

No obstante, la mayoría de estos procesos involucran solo el análisis visual. Las imágenes satelitales siguen siendo subutilizadas debido a varias razones, como la complejidad del manejo de los grandes (y cada vez más crecientes) volúmenes de datos altamente estructurados, así como la falta de experiencia e infraestructura para descargar, almacenar y procesarlos eficiente y eficazmente.

## Panorama internacional

Hace tiempo, las imágenes satelitales solían ser distribuidas en cintas magnéticas. En el 2011, la agencia Geoscience Australia (que custodia los datos geográficos y geológicos en esa nación) transfirió los datos de su territorio de estas cintas a un almacenamiento más moderno: discos mecánicos. Su estrategia fue clave en la iniciativa para *desbloquear el archivo Landsat* de ese país. No tardó en desarrollarse una herramienta diseñada en especí-

fico para facilitar el acceso a estos datos: el *Cubo de Datos Geoespaciales Australiano*.

Para el 2017, esta herramienta tomó un nuevo nombre: *Open Data Cube (ODC)*,<sup>1</sup> tras una serie de evoluciones, como la compatibilidad con distintos sistemas de referencia de coordenadas y formatos de archivos. Su objetivo es permitir enfrentar los retos que se presentan al manejar grandes volúmenes de datos como lo son las imágenes satelitales.

ODC es un proyecto de código abierto que resuelve la necesidad de un mejor manejo de los datos satelitales. La plataforma permite acceder y analizar décadas de imágenes; además, simplifica su uso para observar las condiciones espaciales que captaron, lo cual permitirá que gobiernos, empresas e individuos tomen decisiones más informadas en temas relacionados con el suelo, la vegetación, la erosión costera, la agricultura, la deforestación, los cuerpos de agua o los asentamientos humanos.

A pesar de que el propósito inicial del cubo de datos era aprovechar la riqueza temporal de las imágenes *Landsat*, la flexibilidad de la plataforma permite otra información en el análisis, como: modelos digitales de elevación o cuadrículas de datos geofísicos. Cualquier insumo disponible para el usuario puede ser utilizado en el cubo para enriquecer los resultados, sean datos comerciales, *in situ*, o bien, derivados de un análisis previo.

Varias agencias geoespaciales reconocen que es necesario distanciarse de procesos tradicionales, así como disminuir las barreras causadas por el volumen de los datos y las complejidades relacionadas con la preparación, manejo, almacenamiento y análisis de esta información, para proveer, así, a los usuarios de información para el desarrollo nacional de manera oportuna y, también, mejorar el nivel de detalle en los mapas producidos.

<sup>1</sup> Solución de código abierto para el acceso, manejo y análisis de grandes cantidades de datos de sistemas de información geográfica, sobre todo de observaciones de la Tierra.

Además de Australia, la comunidad de producción de información geoespacial ya se ve favorecida en varias naciones mediante el uso del cubo de datos. Suiza y Colombia son ejemplos de países que ya cuentan con un cubo operacional. Hay, también, ejemplos regionales: en febrero del 2019, la fundación benéfica Leona M. and Harry B. Helmsley destinó 10 millones de dólares a la implementación regional del *Cubo de Datos de África (Digital Earth Africa)*, mientras que el gobierno australiano asignó 10 millones de dólares australianos (alrededor de 7 millones de dólares estadounidenses), además de apoyo técnico y operacional para coordinar esta iniciativa regional.

El INEGI abordó, a su vez, los desafíos mediante la implementación del *Cubo de Datos Geoespaciales de México*. A partir del 2018 se ha invertido en planeación y desarrollo para las ca-

pacidades técnicas y de infraestructura necesaria para atender los retos de *Big Data* propios de esta tarea. A diferencia de los cubos de datos de Suiza o Colombia (ambos funcionan con infraestructura en la nube), el de nuestro país se encuentra en las instalaciones del INEGI, a pesar de que, en términos de volumen, es el mayor de estas tres versiones, derivado de la extensión del territorio mexicano.

## Cubo de datos geoespaciales

Hoy en día, la mayoría de las personas elige ponerse un abrigo o no, cargar con un paraguas o no sin la necesidad de conocer conceptos meteorológicos avanzados, ya que existen varias opciones disponibles de herramientas (integradas en cualquier teléfono celular) que indican las condicio-

Imagen 2

### Panorama global del cubo de datos geoespaciales



Solo 18 meses después de comenzar la iniciativa *Open Data Cube*, había 43 países con cubos de datos operativos en desarrollo o que habían expresado interés en la iniciativa. El interés global es muy diverso y está creciendo muy rápido.

**Fuente:** Killough, Brian. Conferencia. 2018 (DE) [https://www.researchgate.net/figure/18-months-after-starting-the-Open-Data-Cube-initiative-there-are-43-countries-with\\_fig1\\_328995502](https://www.researchgate.net/figure/18-months-after-starting-the-Open-Data-Cube-initiative-there-are-43-countries-with_fig1_328995502)

nes climáticas cada día y, así, permiten tomar decisiones oportunas y evitar contraer un resfriado. De la misma manera, un cubo de datos pretende acercar la información al usuario final en un formato listo para su uso (ver imagen 3).

Esta herramienta es, de manera significativa, más fácil de usar, eficiente y escalable que otros paradigmas de datos; está diseñada para contar con los servicios necesarios para acercar al usuario a la explotación de la información proveniente de las imágenes satelitales. Además, está alineada a los principios de *Big Data*, que ofrecen la posibilidad del procesamiento masivo, escalable y distribuido permitiendo, también, aprovechar técnicas de inteligencia artificial.

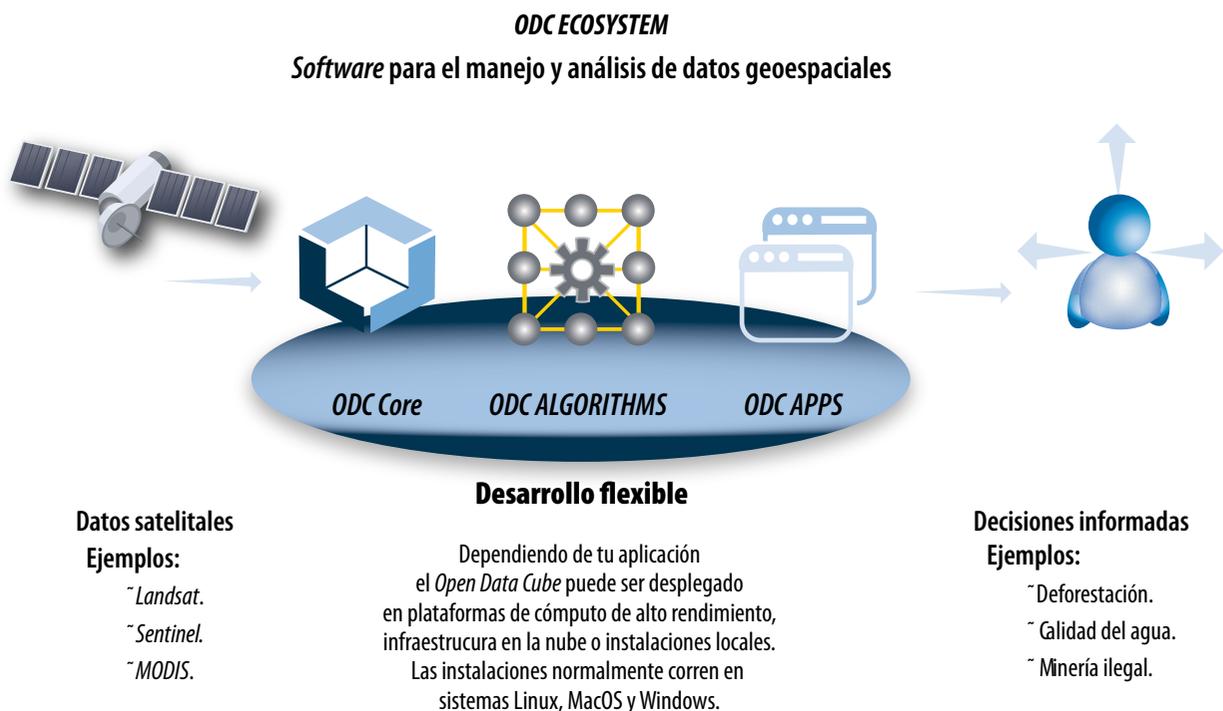
### Arreglo masivo de datos *raster* listos para el análisis

Un cubo de datos es una solución para gestionar ordenada y eficientemente datos espacio-tem-

porales para su análisis; llevado al ámbito práctico, se define como un arreglo masivo de datos *raster*; estos, como las imágenes satelitales, son representaciones espaciales de información en formato matricial: cada celda de esta matriz (también llamada pixel) se puede orientar mediante coordenadas y tiene un valor asignado. Pueden existir varios canales en el mismo pixel de una imagen satelital; esto significa que, así como una foto tomada por la cámara de un teléfono celular suele generarse en el formato común —conocido como RGB, que tiene tres canales (R: rojo, G: verde y B: azul)—, las imágenes satelitales tienen, por lo regular, más canales gracias a sus complejos sensores; en el cubo de datos tienen seis.

Los datos *raster* de un cubo de datos se almacenarán en una sola estructura multidimensional, descrita por varios ejes; las coordenadas en estos son el mecanismo que permite acceder a los datos de forma inequívoca según el espacio, tiempo o canal deseado. Quizá la característica medular de un cubo es que, al integrar la información, este sis-

Imagen 3



Fuente: OpenDataCube.org

tema de coordenadas se homologa para todos los puntos y la ubicación de los píxeles se vuelve independiente de la imagen que lo generó. Así es como facilita el acceso, de entre miles de imágenes de una región o país a través del tiempo, a los datos del área y tiempo especificados por el usuario para su uso inmediato sin verse, además, obstaculizado por su volumen y complejidad.

Si bien la matriz es multidimensional, se le llama cubo de datos por la tridimensionalidad que evoca el nombre de un volumen, ya que, de forma tradicional, para acceder a datos espaciales se usan coordenadas  $(x, y)$ , y al apilar imágenes de la misma zona captadas en distintas fechas se puede observar un punto a lo largo de un periodo, lo cual añade una tercera dimensión, que es el tiempo  $(t)$  (ver imagen 4).

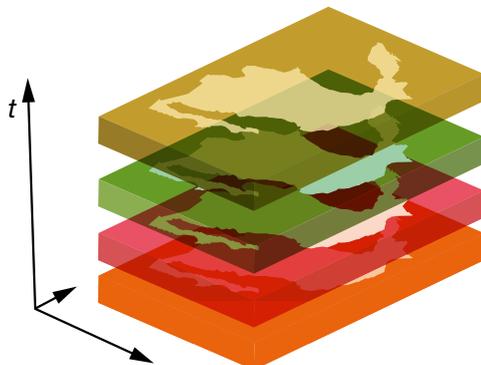
## Datos *raster* listos para el análisis

La arquitectura ordenada (descrita en la sección anterior) favorece un análisis eficiente. Por otro lado, la previa generación de datos estandarizados también es crucial durante la fase de construcción de un cubo.

Las imágenes satelitales ideales para esta herramienta requieren ser preprocesadas hasta lograr un alto nivel de calidad. Estos datos se conocen como *listos para el análisis* (*Analysis Ready Data* o *ARD*) (ver imagen 5). Este término implica que el valor de cada píxel representa el mismo espacio y la misma respuesta espectral independientemente del sensor, el lugar y la época en que fueron captadas, estandarizándolas así para su uso estadístico y analítico a través de series de

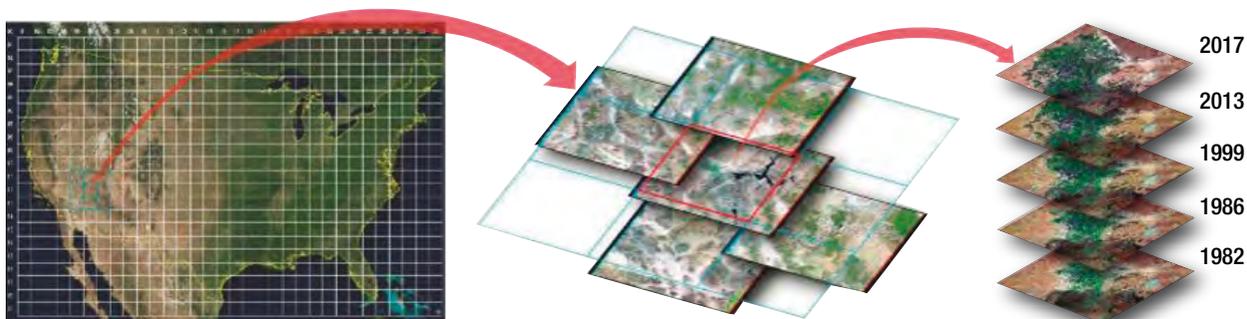
Imagen 4

### Arreglo masivo de datos multidimensionales (*raster*)



Fuente: INEGI.

Imagen 5



Fuente: U.S. Landsat Analysis Ready Data (DE) <https://www.usgs.gov/>

tiempo. Además, deben cumplir con descripción de metadatos y calibraciones radiométrica, geométrica, solar y atmosférica.

Generar estas imágenes requiere de un esfuerzo computacional adicional aplicado a las comúnmente disponibles en los repositorios (datos crudos). Aunque por lo general es posible solicitar apoyo de forma directa a las agencias correspondientes, el acceso al material con estas características se irá incrementando. En la actualidad, agencias y proveedores trabajan para su pronta disponibilidad en la nube; por ejemplo, se espera que las imágenes *Landsat ARD* sean de fácil acceso de esta forma en cuestión de meses.

Podría compararse, entonces, un cubo con un video: varios cuadros (*frames*) con datos ordenados de manera temporal; para el video serían tal vez datos RGB, mientras que para una imagen satelital se agregarían las otras bandas no visibles, como el infrarrojo. Siendo aún más específicos, este sería un video en cámara rápida (*Time-Lapse*) y las correcciones de parpadeo o movimiento de la cámara (*flicker, motion control*) equivaldrían al procesamiento que genera los datos *ARD* ordenados, ya que alinean los cuadros de forma correcta y remueven efectos indeseados por variaciones en la luz.

Las imágenes *ARD* en un cubo de datos serán transformadas a una proyección geográfica definida y, así, esta organización permitirá su análisis inmediato sin que sea necesaria ninguna acción adicional por parte del usuario. De esta forma, la adopción del cubo impulsará los procesos involucrados en la explotación de imágenes satelitales en el INEGI: ahora, los datos se podrán almacenar, administrar, acceder y analizar de manera más eficiente.

## **Cubo de Datos Geospaciales de México**

El 4 de marzo de 2019, el INEGI recibió el archivo histórico *Landsat ARD* con más de 109 mil imágenes, que fueron generadas a partir de su versión original y luego enviadas por el personal de USGS/NASA; su volumen en formato comprimido ascien-

de a 30 TB. Este archivo ya está cargado y respaldado en la infraestructura del Instituto y su volumen es de 90 TB cuando se descomprime. Descargarlo y procesarlo —para un país del tamaño de México— desde un repositorio en línea resultaría computacionalmente demandante y poco eficiente.

Otro paso hacia la implementación del *Cubo de Datos Geospaciales de México* fue la definición de una proyección adecuada para el archivo *Landsat ARD*. El INEGI favoreció la proyección cónica equiárea de Albers que, como lo implica el nombre, proyecta las áreas de manera proporcional. Esto facilitará las estadísticas de superficie, que es un enfoque común para aquellos indicadores de los Objetivos de Desarrollo Sostenible (ODS) —de la *Agenda 2030*— basados en información geoespacial.

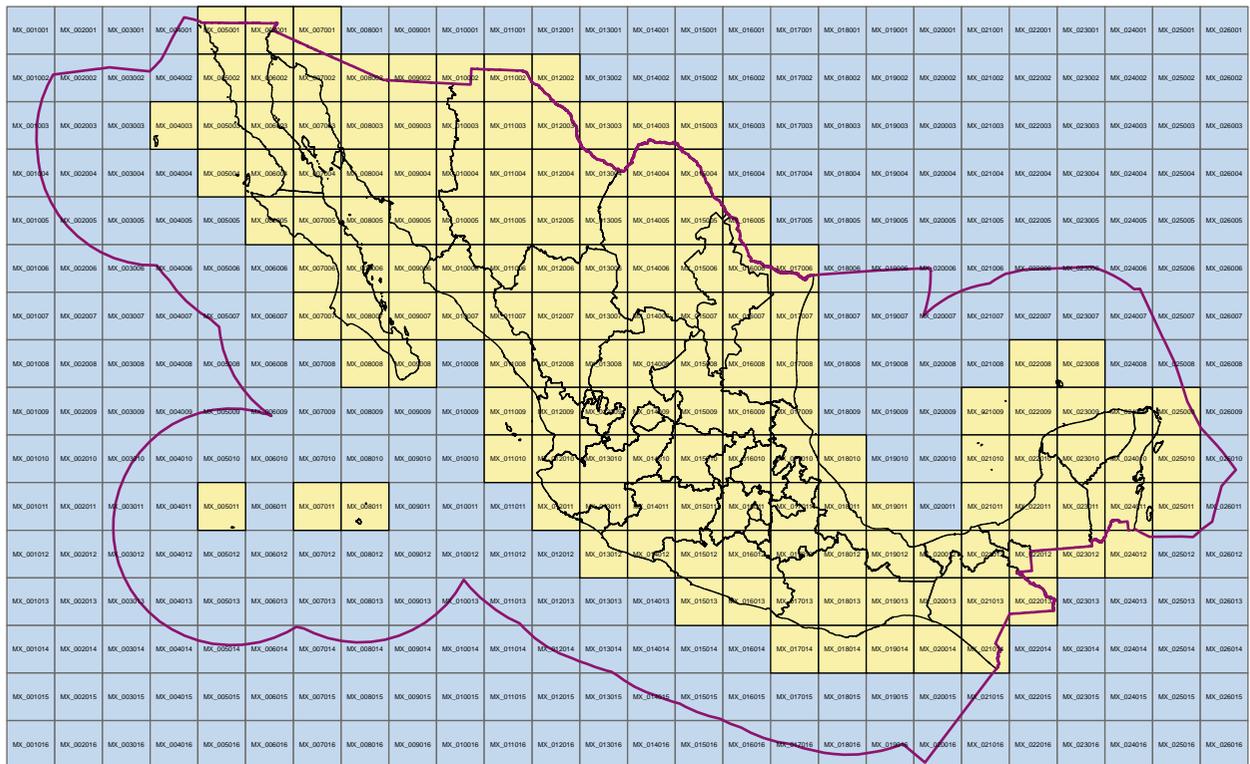
Sin embargo, una matriz de datos con toda esa información resultaría en un archivo difícil de manejar de manera tradicional, en especial si se desea realizar un análisis a nivel nacional. Considerando píxeles de 30 m por lado, provenientes de imágenes *Landsat*, sería como editar un video con resolución de alrededor de 105 mil píxeles de ancho y 70 mil de alto: el formato de ultra alta definición 4K tiene apenas 3 840 por 2 160.

Para facilitar esta y otras tareas computacionalmente complejas, la proyección de las imágenes se subdivide en mosaicos de 5 mil píxeles por lado. Este nuevo nivel de abstracción lógica en la arquitectura de los datos los distribuirá en archivos idénticos en su estructura con la finalidad de *paralelizar* los procesamientos computacionales altamente demandantes (ver imagen 6). El diseño de estos mosaicos de 5 mil píxeles por lado equivale a 150 km, lo que garantiza, además, compatibilidad para otras resoluciones como píxeles de 10 m por lado, en el caso de datos generados por el sensor de *Sentinel-2*.

## **Ventajas del Cubo de Datos Geospaciales de México**

Las metodologías tradicionales, en las que se incluye el uso y análisis de imágenes satelitales, conllevan

Imagen 6



Fuente: INEGI.

van un manejo tradicional y gran dependencia en la fotointerpretación (identificación de rasgos mediante análisis visual). Las dos grandes ventajas del cubo mexicano son:

- Aprovechamiento de recursos tecnológicos. Utilizando una instancia de un cubo de datos, en lugar de que cada imagen se descargue, almacene y analice en las computadoras de los especialistas, se pueden aprovechar los recursos tecnológicos y los tiempos destinados a los preparativos de los procesos. Con el análisis de datos efectuado mediante el archivo previamente generado, con imágenes alineadas, ordenadas y estandarizadas, se permitirá automatizar algunas fases de las metodologías, como la descarga y distribución de las imágenes a los especialistas. Por otro lado, la precisión espacial es mucho mayor.

Asimismo, al permitir procesamientos masivos, esta tecnología facilitará aplicar algoritmos de inteligencia artificial a grandes volúmenes de datos. Utilizando este atributo, la inspección visual también podría ser asistida por clasificaciones supervisadas a nivel nacional efectuadas de manera previa. De esta forma, se puede dirigir la atención de los pocos especialistas que deben analizar toda la extensión de un país a la inspección de regiones con características demasiado confusas para los algoritmos y, así, reducir los considerables tiempos que demandan la exhaustividad de la inspección y análisis visual.

- Aprovechamiento de los datos. Las imágenes con nubosidad no serán descartadas en su totalidad, como es común, ya que ahora se pueden eliminar solo los pixeles afectados por las nubes; por ejemplo, para construir imágenes

continuas y libres de ellas para el monitoreo del crecimiento urbano (ver imagen 7).

Además de las posibles mejoras a los procesos actuales, otra ventaja de contar con un arreglo estructurado de imágenes alineadas correctamente, como es el cubo de datos, es un mayor aprovechamiento de las imágenes satelitales y la libertad de selección de las áreas y periodos específicos de estudio que mejorarán, por ejemplo, el monitoreo de cambios en los recursos naturales a lo largo de décadas. Para ilustrar esta idea, en la imagen 8 se aprecia el cambio en la vegetación a lo largo de casi 30 años en una región (el rojo representa vegetación abundante). Estas posibilidades pueden permitir, sin duda, ampliar el catálogo de productos de información geoespacial.

Aunado a lo anterior, el *Cubo de Datos* está desarrollado de inicio a fin en el lenguaje de programación *Python*, el cual tiene una curva de aprendizaje

muy suave y, por lo mismo, es de gran popularidad en las comunidades de ciencia de datos y de procesamiento de imagen. Tras la construcción de un cubo de datos, el usuario final (que tal vez sea un especialista temático) solo debería preocuparse por dominar este lenguaje de programación de forma general y luego, las instrucciones básicas propias de la biblioteca (*set* de comandos) del cubo de datos para el manejo de la información. Esta y muchas otras bibliotecas prediseñadas se encuentran disponibles de forma pública y gratuita (OpenDataCube.org).

## Análisis temporal a nivel pixel

El *Cubo de Datos* cuenta con una serie de algoritmos (biblioteca *Data Cube Stats*) que generan varios análisis directos, los cuales han sido aprovechados en varios temas.

Los productos se producen para una zona de interés en un periodo deseado mediante recortes o rebanadas de la matriz utilizando las coordenadas correspondientes. Se basan en el análisis de series de tiempo; las operaciones usan la serie independiente y propia de cada pixel en la zona seleccionada; en consecuencia, todos sus resultados se presentan, también, a nivel pixel.

Imagen 7

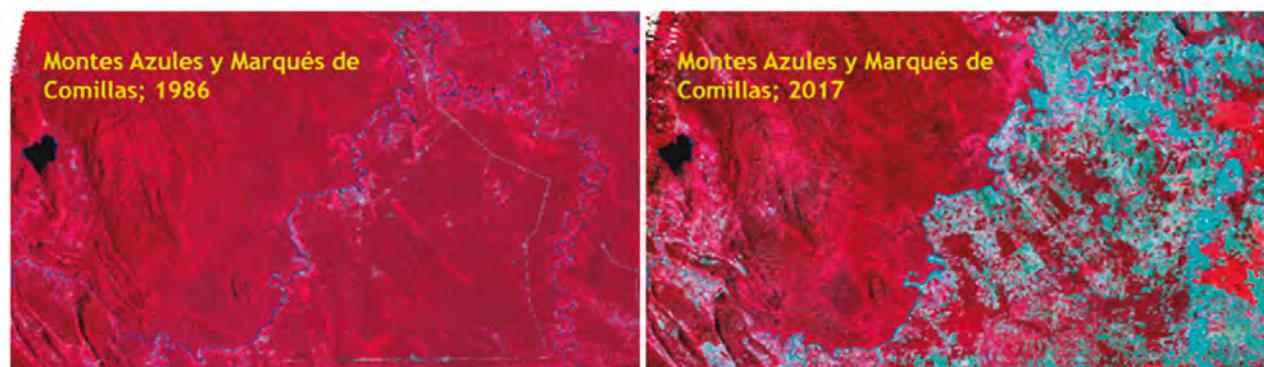
### Secuencia de imágenes resumen en Mérida, Yuc., generadas con los datos *Landsat* del 2000, 2005, 2010 y 2017 mediante el *Cubo de Datos*



Fuente: INEGI.

Imagen 8

### GEO-XV Plenary Session 2: EO in Support of the SDG, Kyoto, Japan, October, 2018



Fuente: INEGI.

Esta es una importante contribución que complementa e impulsa el uso del *Cubo*, ya que hace disponibles, de manera inmediata, herramientas de uso común en imágenes satelitales para generar estadística, como los índices de diferencia normalizada, por ejemplo, el Índice de Áreas Construidas de Diferencia Normalizada (NDBI, por sus siglas en inglés), que es una frecuente vía de análisis territorial en estudios urbanísticos e infraestructuras y en la comparación de la evolución del desarrollo urbano en el tiempo, ya que resalta las zonas con superficies edificadas o en desarrollo de construcción frente a las habituales áreas con vegetación o desnudas.

Quizá, una de las técnicas más comunes en el análisis de imágenes satelitales es el Índice de Vegetación de Diferencia Normalizada (NDVI, por sus siglas en inglés), que resalta la flora del resto de los elementos presentes en una imagen.

Estas versiones de índices de diferencia normalizada generadas con el *Cubo de Datos* calculan, además, los valores estadísticos (máximo, mínimo y media), lo cual permite aprovechar la información aún más, ya que estos describen el comportamiento del índice a lo largo del tiempo (en este caso, un año), lo que proporciona una mejor idea de la respuesta (ver imagen 9); por ejemplo, en el caso del NDVI, conocer el comportamiento de la vegetación durante un año permite vincular variables meteorológicas a las manifestaciones

temporales periódicas o estacionales de las plantas, lo cual favorece una mejor clasificación para cada pixel, a diferencia de lo que se podría concluir considerando solo el valor observado en una fecha determinada.

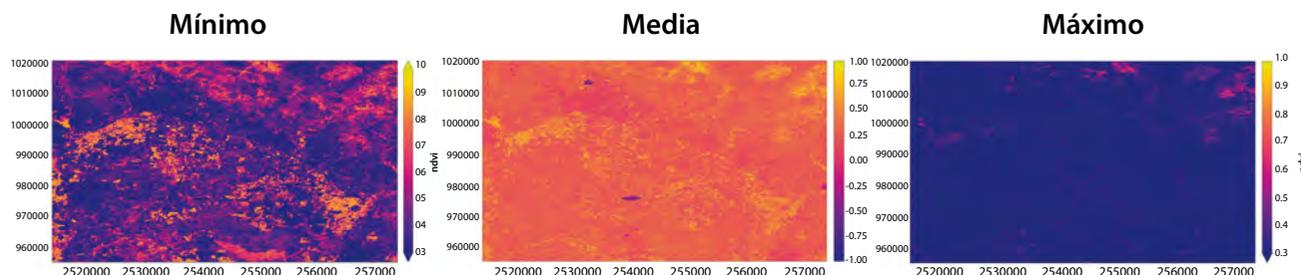
### Geomediana y observaciones de agua desde el espacio

A esta colección de productos basados en el análisis a nivel pixel se agregan estos dos importantes desarrollos diseñados para el mayor aprovechamiento del *Cubo de Datos*.

La geomediana es un algoritmo utilizado para realizar un resumen estadístico multivariado para todos los valores observados de un mismo pixel en un periodo indicado. Aplicándolo a una región, se produce una imagen compuesta de *pixeles resumen*, la cual mantiene consistencia espacial, incluso en los límites entre escenas; típicamente, se obtienen imágenes continuas y libres de nubes (ya que se descartan las observaciones de pixeles donde hay nubosidad). El algoritmo trabaja con todas las bandas del pixel a la vez por lo que, además, se conserva la razón entre estos valores; debido a ello, es posible utilizar el producto de estas operaciones como el insumo de otros procesos de uso frecuente que parten de estas relaciones entre las bandas, como los presentados en la sección anterior (NDBI y NDVI).

Imagen 9

### NVDI en Guanajuato, 2015



Fuente: INEGI.

La geomédiana nacional 2015 fusiona las 138 captaciones realizadas por los satélites *Landsat* durante todo ese año, que cubren el territorio continental e islas de México; cada una representa regiones de 182 km por 185 km, aproximadamente, y es captada cada ocho días. En otras palabras, esta imagen continua y libre de nubes resume los 7.5 TB de datos que son 6 074 imágenes *Landsat*, sin perder nivel de detalle espacial (ver imagen 10).

Por otra parte, una aplicación más del *Cubo* es un algoritmo para detectar y monitorear la presencia de agua mediante una clasificación supervisada (*agua o no agua*). Este se llama *Observaciones de agua desde el espacio o Water Observations from Space (WOfS)* y calcula el porcentaje de presencia de este líquido en el periodo seleccionado (ver imagen 11).

## Integración de otras fuentes para realizar clasificaciones

Además de simplificar el análisis de un gran conjunto de imágenes *Landsat*, los cubos de datos geospaciales pueden complementar metodologías que implican el uso de otras fuentes de información, ya sea tipo *raster* (como un modelo digital de elevación) o vectorial (como los datos de campo).

Un ejemplo de esto es el ejercicio realizado para la identificación de cultivos y cálculo de su superficie utilizando técnicas de aprendizaje automático. En una zona del estado de Guanajuato, previamente definida, se aplicó el clasificador supervisado *Random Forest*. El cultivo de interés para este ejercicio fue el de sorgo, principal producto de la zona estudiada.

Imagen 10

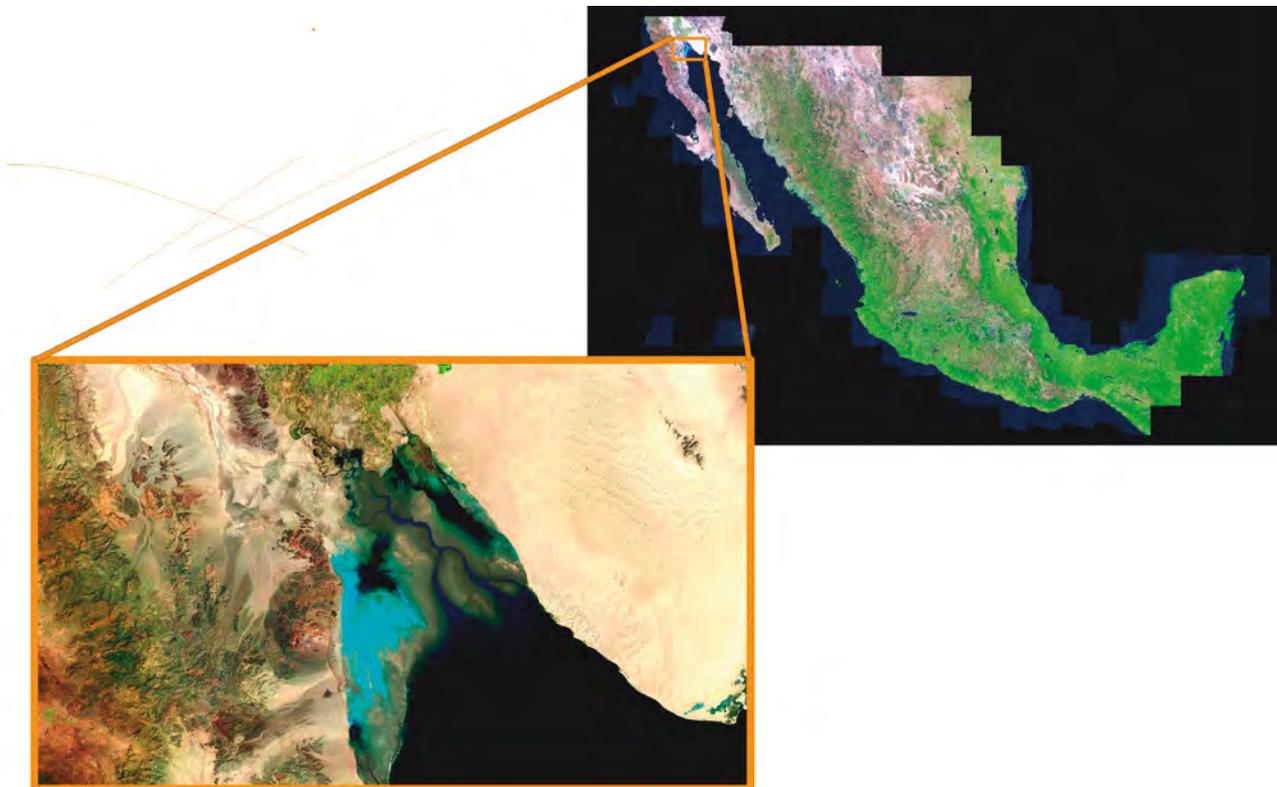


Imagen continua y libre de nubes del territorio mexicano generada con 6 074 imágenes *Landsat* del 2015 mediante el algoritmo mediana geométrica y acercamiento al delta del río Colorado, mar de Cortés.

Fuente: INEGI.

El insumo obtenido de manera directa del *Cubo de Datos Geoespaciales de México* fueron imágenes *Landsat* correspondientes a junio, julio y agosto del 2015; el periodo se eligió a partir del ciclo de cultivo del sorgo, pues en junio la planta alcanza a cubrir el suelo y en agosto se encuentra en su etapa de maduración.

De manera adicional, se utilizó información de campo correspondiente a los terrenos con cultivos en la zona de estudio durante el mismo año. Debido a la resolución espacial de *Landsat* de 30 m por

pixel, se consideraron solo los terrenos de superficie mayor (ver imagen 12).

Los resultados fueron satisfactorios en los polígonos de mayores áreas. Sin embargo, se presentó confusión para los más pequeños; por este motivo, las pruebas se están recreando en la actualidad utilizando imágenes de mayor resolución espacial y temporal (*Sentinel-2*).

También, se realizó el ejercicio usando una imagen satelital de alta resolución (6.5 m) y se com-

Imagen 11

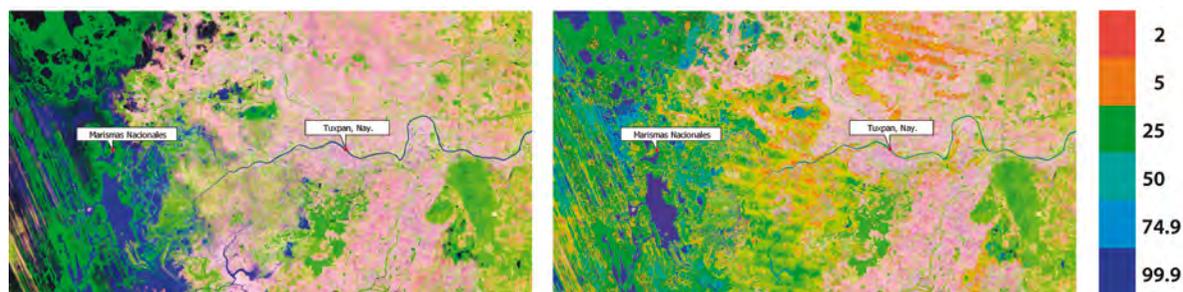
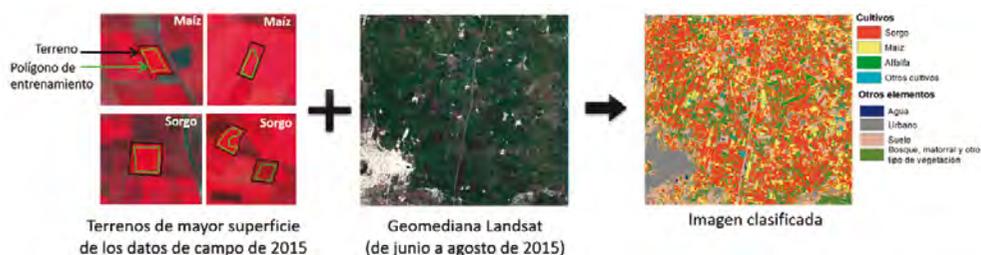


Imagen simple y sobreposición de una representación en color del porcentaje de agua presente durante el 2015 en las regiones de Marismas Nacionales y Tuxpan, Nayarit, mediante el algoritmo *WOFI*.

Fuente: INEGI.

Imagen 12

### Proceso para realizar clasificación supervisada de cultivos utilizando datos de campo y *Landsat* de junio a agosto del 2015 para entrenamiento del algoritmo *Random Forest*



Fuente: INEGI.

pararon las áreas resultantes de los cultivos clasificados de un municipio con cobertura completa en ambas imágenes (Valle de Santiago). Las cifras resultantes son consistentes entre sí; el cuadro de la imagen 13 muestra los resultados para superficie cultivada del cultivo principal del estudio (sorgo) y de maíz, según el insumo utilizado en el ejercicio.

## Conclusiones

Con el *Cubo de Datos Geoespaciales de México* hay un cambio de paradigma en cómo el INEGI entrega información a los usuarios. Históricamente, se han presentado productos finalizados, con tabulados básicos y representatividad nacional, estatal y municipal. En este caso, se pondrían a disposición del usuario las imágenes y los algoritmos con los cuales puede realizar sus propias estimaciones.

Con cada año que pasa, los nuevos satélites están creando volúmenes de datos cada vez más significativos con cobertura global. La falta de información ya no es un factor limitante, pero

almacenar y procesar esta cantidad de datos representa un nuevo reto para el desarrollo de capacidades. No obstante, hay confianza en que el enfoque basado en generar la infraestructura necesaria para el uso de los datos de observaciones de la Tierra ofrece un gran potencial para reducir las barreras técnicas para explotarlos en todo su potencial. Con una gran y creciente comunidad internacional de usuarios y desarrolladores, las aplicaciones del *Cubo de Datos* son virtualmente infinitas.

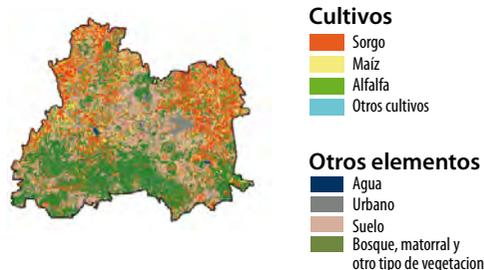
La herramienta ofrece excelentes soluciones para los problemas principales de los usuarios, como acceso, preprocesamiento y análisis eficientes para respaldar las aplicaciones de los usuarios, además de que permite minimizar el tiempo y el conocimiento especializado requerido para acceder y preparar los datos satelitales. En consecuencia, con estos cambios en el INEGI se prevé un gran éxito e impacto a escalas local, regional y mundial, apoyando agendas locales y globales prioritarias, como las que se encuentran en los ODS de la Organización de las Naciones Unidas, el Marco de Sendai y el Acuerdo de París, entre otros.

Imagen 13

### Comparación de área cultivada estimada de sorgo y maíz en el municipio de Valle de Santiago, 2015

| Fuente y resolución                                    | Área de sorgo (ha) | Área de maíz (ha) |
|--|--------------------|-------------------|
| Varias imágenes de satélite <i>Landsat</i> 2015 (30 m) | 17 420             | 7 469             |
| Una imagen de satélite <i>RapidEye</i> 2015 (6.5 m)    | 18 858             | 5 315             |

#### Clasificación de cultivos del municipio de Valle de Santiago



Fuente: INEGI.

Uno de los retos más relevantes al que nos enfrentamos los promotores del uso de la información geoespacial es lograr la necesidad de hacer que la dimensión geoespacial sea realmente transversal. Para ello, el INEGI cuenta con el beneficio de tener geografía y estadística en una sola institución nacional, lo que ha permitido a México una mejor integración y uso de sistemas de información complementarios. Con las herramientas asociadas de esta integración es posible georreferenciar estadísticas relevantes, así como determinar la ubicación de las desigualdades económicas y sociales, además de los riesgos y daños causados por desastres naturales. En este aspecto, el *Cubo de Datos* representa, también, una importante posibilidad para hacer eficiente la integración y el análisis de datos geográficos y estadísticos para un mejor diseño y monitoreo de las políticas públicas.

---

## Fuentes

Roberts, D.; N. Mueller & A. McIntyre. "High-Dimensional Pixel Composites From Earth Observation Time Series", in: *IEEE Transactions on Geoscience and Remote Sensing*. 99, 2017, 1-11.

Killough, Brian. *Overview of the Open Data Cube Initiative*. 2018, 8629-8632. 10.1109/IGARSS.2018.8517694.

Lewis, A. *et al.* "The Australian Geoscience Data Cube-Foundations and Lessons Learned", in: *Remote Sensing of Environment*. 2017.

Piotrowicz, Luke. "The 40th Anniversary of Landsat, Australia's 33 year archive of Landsat data", in: *AusGeo News*. September 2012, Issue No. 107.

Mueller, N.; A. Lewis, D. Roberts; S. Ring; R. Melrose; J. Sixsmith; L. Lymburner; A. McIntyre; P. Tan; S. Curnow & A. Ip. "Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia", in: *Remote Sensing of the Environment*. 174, 2016, 341-352.

Picoli, Michelle *et al.* "Big earth observation time series analysis for monitoring Brazilian agriculture", in: *ISPRS Journal of Photogrammetry and Remote Sensing*. 2018. DOI: 10.1016/j.isprsjprs.2018.08.007.

Roberts, Dale; Bex Dunn & Norman Mueller. *Open Data Cube Products Using High-Dimensional Statistics of Time Series*. 2018, 8647-8650. 10.1109/IGARSS.2018.8518312.

Strobl, Peter; Peter Baumann; Adam Lewis; Zoltan Szantoi; Brian Killough; Matthew Purss; Max Craglia; Stefano Nativi; Alex Held & Trevor Dhu. "The Six Faces of the Data Cube", in: *Conference: Big Data from Space*. Toulouse, France, 2017.