

# La calidad en los datos espaciales

LUIS GERARDO ESPARZA RÍOS

HÉCTOR MARIO GÓMORA MORALES

El desarrollo de la sociedad se basa en el conocimiento de sus características, sus recursos naturales, sus problemas y sus necesidades. Es lo que permite identificar prioridades y diseñar alternativas para el futuro. En este proceso las sociedades tienen que renovar conceptos, superar ideas o formas de hacer las cosas que antes se creían inmutables: se encuentran ante al reto de cambiar paradigmas, ante la disyuntiva de renovarse o estancarse.

La obtención de los datos espaciales se enfrenta actualmente con esta situación. En su generación intervienen factores que les dan diferentes niveles de confiabilidad y éstos deben ser conocidos. Ello permite avanzar en el estudio de los problemas incluidos en su recolección y procesamiento; y sobre todo, reconocer que los datos, en su carácter de producto dirigido a la gente, deben ser evaluados para que cada usuario potencial estime su utilidad. Se trata de adoptar el enfoque de calidad que tiene cada vez más difusión entre los productores y usuarios de los diversos sistemas de análisis geográfico.

## La calidad: historia y fundamentos

Podemos definir la calidad como la totalidad de las características de un producto o servicio que le confieren su aptitud para satisfacer necesidades explícitas o implícitas.<sup>1</sup> Las necesidades explícitas son aquellas declaradas por el fabricante y constituyen el atractivo del producto: los programas en el caso de una computadora o, para la cartografía, la actualidad de los datos representados. Las necesidades implícitas son las que no se declaran por obvias, dependiendo del tipo de producto: los programas de una computadora deben ser amigables en su manejo, un mapa debe ser entendido con facilidad.

Desde que la calidad fue asociada con la exigencia de que el producto cumpliera con requisitos predeterminados se estableció la relación entre este concepto y el de normalización, que es la elabora-

ción, difusión y aplicación de normas encaminadas a uniformar determinadas actividades. Una característica importante de este proceso es que surge por necesidad; es decir, se crean normas cuando se tienen dificultades derivadas de la falta de homogeneidad en los productos. Por ello Ariza<sup>2</sup> resalta que es la forma de establecer soluciones productivas a situaciones repetitivas.

Ahora bien, aunque el concepto de calidad se originó en la producción de bienes, sus principios básicos son aplicables a otras actividades, como la generación de datos espaciales.

La elaboración de mapas fue durante mucho tiempo el producto por excelencia del quehacer geográfico y también aquí se establecieron especificaciones para unificar los productos. La normatividad básica se concentraba en aspectos como los sistemas de referencia, la distribución y clasificación de hojas (en el caso de las series cartográficas), el formato de presentación y la simbología. Al igual que sucedió con la generación de bienes industriales, la atención originalmente estaba enfocada en el mapa.

Desde mediados del siglo xx esta situación empezó a cambiar debido a varios factores, entre ellos, el desarrollo de tendencias cuantitativas en el análisis geográfico, la tecnología de cómputo, la cartografía automatizada y los sistemas de información geográfica. Más recientemente, en los noventa llegaron las bases de datos espaciales y la tecnología de internet, que posibilitaron la transferencia e intercambio a distancia de los datos.

Este medio siglo de avances tecnológicos trajo un cambio importante en los procesos cartográficos. No sólo fue posible "obtener los mismos productos de antaño con notables ventajas y facilidades",<sup>3</sup> sino que se creó otra vertiente en el quehacer geográfico: estudiar el proceso de recolección y tratamiento de los datos que sustentaban las representaciones cartográficas y la forma en que esto repercute en su confiabilidad.



Estas consideraciones llevaron a tomar conciencia de la necesidad de que los datos espaciales contaran con sus propios principios de calidad. Algunos de éstos son similares a los de la industria; otros fueron desarrollados según las características propias de esta clase de datos.

### El dato espacial y su calidad

Dado que los datos son el inicio de la cadena que lleva hacia el conocimiento y la toma de decisiones,<sup>4</sup> es en ellos donde se debe poner mayor atención.

¿Qué se debe entender por dato espacial? En el espacio geográfico existen objetos naturales y artificiales (lagos, ríos, carreteras, localidades) y cada uno de ellos puede ser visto como una entidad susceptible de ser representada con un registro único en una base de datos. Dicho registro digital es el *dato espacial*. Entre sus características hay dos relevantes: la primera es su localización mediante una referencia en tres dimensiones (latitud, longitud y altitud); la segunda está formada por dos componentes: 1) gráfico: es la imagen digital que representa al objeto; 2) descriptivo: tabla de valores con las características del dato: si es carretera, su número de carriles y su jurisdicción; si es localidad, su nombre, número de habitantes, etc. Es decir, el dato espacial conjunta una imagen susceptible de ser ubicada en la cartografía con un registro que nos permite saber algunas cosas sobre el objeto que representa la imagen.

Debido a esta doble naturaleza, el dato espacial es un concepto más complejo que los simples datos numéricos con los que estamos familiarizados, y esta complejidad tiene como consecuencia que está expuesto a diferentes tipos de error. Conviene mencionar dos ejemplos. Por un lado, ubicar un dato espacial –un río o una localidad– y determinar las coordenadas que corresponden a su ubicación puede hacerse directamente sobre el terreno, pero por lo general se utilizan fuentes indirectas como fotos aéreas y mapas, y en tales casos la ubicación de coordenadas es menos exacta. Por otro lado, los datos espaciales pueden ser muy dinámicos y cambiar rápidamente, como es el caso del uso del suelo, donde las áreas de cubierta vegetal natural o el avance de la mancha urbana cambian a veces en cuestión de días. En circunstancias así, la actualidad de los datos es muy breve.

En consecuencia, el proceso de obtención, registro y manejo de los datos espaciales está expuesto a diferentes márgenes de error que difícilmente serán eliminados del todo, aunque pueden ser reducidos y hasta cierto punto controlados. Se llega así a una idea básica: *calidad de datos espaciales no significa la ausencia de errores ni la existencia de una información perfecta*.

Entonces, ¿cuál es la idea de la calidad en esta clase de datos? Principalmente, que el productor determine con base en pruebas específicas, los niveles de exactitud de sus datos y los dé a conocer a los usuarios potenciales. Esto se hace evaluando diversas propiedades, porque la calidad de un producto es un concepto que integra varias de sus características. ¿Cuáles son esas características o parámetros mediante los cuales se evalúa la calidad? Para un auto pueden ser el rendimiento por litro de combustible, la comodidad de interiores o la diversidad de accesorios; para un aparato electrónico, el diseño, las funciones de que dispone o la facilidad de manejo; en los datos espaciales normalmente se desea actualidad, veracidad y exactitud en la ubicación, entre otros.

La calidad de los datos espaciales debe ser expresada usando diversos parámetros. Éstos son los *criterios de calidad de los datos espaciales*.

### Criterios de calidad de datos espaciales

Son de dos grandes tipos: *cuantitativos* y *cualitativos*. Entre las organizaciones dedicadas a estudiar la calidad no existe consenso sobre cuáles criterios forman cada grupo. A continuación se mencionan aquellos donde hay más coincidencias.

Los *criterios cuantitativos* miden el grado en que el conjunto de datos cumple con sus especificaciones. La concepción de éstos se deriva de cada una de sus características.

El dato espacial tiene una posición expresada en coordenadas y, como toda medida, está expuesto a un cierto grado de error que debe ser conocido. En otras palabras, qué tanto se acercan los valores de las coordenadas a los verdaderos o a los que se obtienen de otra fuente que ha sido considerada de mayor confiabilidad. Esto se conoce como *exactitud posicional*.

Además, el dato espacial tiene ligados atributos que permiten conocer ciertas cosas de él. En el caso



de una carretera, si es de tipo federal, si tiene cuatro carriles, etc. La recolección e incorporación a la base de datos de estos atributos también se expone a errores y es conveniente conocer su nivel. La *exactitud temática* corresponde a la evaluación de la veracidad de los datos en relación con la realidad o con otra fuente considerada confiable.

Es necesario comprobar la veracidad y congruencia de las referencias temporales, verificar que las fechas registradas sean correctas y si existe una secuencia lógica entre ellas, ya que no es congruente, por ejemplo, que un mapa indique como año de impresión 1990 y, además, que fue elaborado con análisis de fotografías aéreas de 2001. Esto es lo que estudia el criterio de *exactitud temporal*.

Normalmente los datos espaciales no se integran aislados, sino en conjuntos que reúnen varios tipos: ríos, cuerpos de agua, caminos, tipos de cubierta vegetal, por citar algunos. En este nivel interesa que estén todos los elementos que deben figurar y evitar los sobrantes. Por ejemplo, si se trata de consignar sólo los caminos pavimentados deben estar todos ellos, sin terracerías ni veredas. La idea corresponde a verificar la integridad de los componentes, su *completitud*.

En la *consistencia lógica* de la construcción de la base de los datos espaciales se establece un conjunto de normas informáticas referido comúnmente como el *modelo de datos espaciales*, que reglamenta diversos aspectos, como la geometría de los objetos por representar, las formas en que éstos pueden tocarse o superponerse gráficamente, los valores que pueden tener los atributos, etcétera.

Para medir cada uno de estos criterios cuantitativos existen diferentes métodos de prueba y formas de expresarlos numéricamente, como el porcentaje de errores u otros de tipo estadístico más complejos como la desviación estándar o el error medio cuadrático. Además, lo ideal sería que todos estos criterios fueran validados directamente en el terreno, que es la fuente más confiable, pero debido a la complejidad para verificar de esta manera todos los datos espaciales (por ejemplo, todas las carreteras o los ríos) normalmente se trabaja sobre una muestra y se recurre a fuentes (mapas, imágenes) que, por su fecha, detalle u otras características han sido consideradas más confiables.

Los criterios *cualitativos* contienen información general; aportan elementos para conocer más deta-

lladamente los datos y evaluar la posibilidad de usarlos según las necesidades de un determinado usuario.

Un aspecto importante es explicar los motivos que dieron lugar a la creación del conjunto de datos; es decir, esbozar el propósito para el cual fue creado. Esto corresponde al criterio llamado *objetivo*.

El *linaje, genealogía o historia de los datos* permite formarse una idea adecuada del conjunto de datos, ya que el usuario requiere una descripción de los aspectos relevantes considerados en su creación, que resalte las principales etapas de su existencia, y las actualizaciones o las reestructuraciones a las que fue sometido.

Otra referencia importante es la utilización institucional o particular de los datos. Esto sirve para los usuarios potenciales que pretendan darles un uso similar.

Estos ocho criterios son la base para desarrollar la metodología de calidad. El productor realiza la fase denominada *evaluación*, en la cual aplica pruebas para determinar los valores o "calificaciones" que sus datos obtienen en los criterios cuantitativos. Asimismo elabora la información de referencia que corresponde a los criterios cualitativos. Luego da a conocer ambas cosas, con lo cual realiza la *declaración de la calidad*. Es esto lo que el usuario potencial consulta, concentrándose evidentemente en los aspectos que a él le interesan.

La declaración de la calidad implica un compromiso para el productor. Si reporta, por ejemplo, que sus datos tienen coordenadas de ubicación cuyo margen de error nunca supera los 10 metros, está garantizando al usuario esa exactitud. Aquí es donde se muestra cómo interactúan las dos vertientes de la calidad: el cumplimiento de las normas y la utilidad del producto. Así, la calidad es una idea que puede variar de un usuario a otro dependiendo de sus objetivos. Una exactitud como la mencionada antes puede ser adecuada para quien realiza estudios de grandes áreas boscosas y puede permitirse unos cuantos metros de error. En contraste, un levantamiento catastral requiere datos cuyas coordenadas tengan un margen de error mucho menor, del orden de centímetros.

Se trata de una responsabilidad compartida: el productor debe usar los criterios cuantitativos y cualitativos para determinar la calidad de sus datos y declararla. El usuario debe entender tales criterios



para interpretar correctamente los resultados y evaluar sus posibilidades y limitaciones. El hecho de que el productor dé a conocer los niveles de exactitud representa un compromiso para él y también un factor de confianza para quienes los usen, pues así disponen de parámetros para conocer su nivel de confiabilidad.

Como apunta Mas Mayoral: “no basta con establecer una norma o especificaciones de la información geográfica que se integra en un sistema de información, hay que asegurar la calidad de la misma durante todo su proceso de producción y tratamiento para que las especificaciones de la información no queden en meras normas teóricas”.<sup>5</sup> Es decir, la existencia de normas técnicas es indispensable, pero debe ser complementada con la información que el productor proporcione al usuario potencial para que conozca el nivel de cumplimiento de calidad del producto.

Los datos espaciales no tienen calidad por el solo hecho de ser digitales. Si se tiene conciencia de ello se entenderá que la calidad significa conocer los niveles de cumplimiento de las normas y así tener los elementos para determinar si son adecuados para una determinada aplicación.

### Hacia un nuevo paradigma: la calidad como elemento de la IDEMEX

Actualmente el INEGI coordina la *Infraestructura de Datos Espaciales de México* (IDEMEX), que conjunta recursos, normas, tecnologías, políticas, marco legal, administrativo y organizacional para una efectiva creación, recopilación, manejo, acceso, distribución, compartición y uso de los datos espaciales.<sup>6</sup> Dentro de este concepto integral, habrán de concurrir productores y usuarios, conectados por el producto que, a su vez, será elaborado con base en normas técnicas. La metodología de calidad se concibe como el factor que puede articular adecuadamente estos elementos.

Actualmente se trabaja en el diseño de un Modelo de Calidad de Datos Espaciales que en el futuro proporcione las bases para aplicar el enfoque de calidad en el contexto de la IDEMEX. Todo ello con el objetivo de que los datos se difundan más y sean accesibles para una mayor cantidad de usuarios, contribuyendo a cumplir la misión de dar a la sociedad el servicio público de información.

Se dice que la calidad cuesta porque su aplicación requiere recursos; sin embargo, hay que tener presente que los costos de su carencia pueden ser superiores. Recordemos además que los datos espaciales son el soporte de decisiones que tienen trascendencia en diferentes ámbitos, regional o incluso nacional. Por ello, es deseable que el enfoque de calidad sea una política institucional, y que se le dé particular importancia a su difusión.

En esto consiste el cambio de paradigma: la calidad debe ser vista como parte integral y regular del proceso de generación de datos espaciales, ya que favorece la existencia de datos con una estructura común, concebidos con el principio de “elaborarlo una vez y usarlo muchas veces”, con ahorro de recursos. La aplicación de la metodología de calidad responde a la intención de crear un lenguaje común para la generación de los datos espaciales; y la misma historia de la calidad muestra que la normalización es indispensable para un trabajo cuyos resultados sean útiles para la sociedad.

El INEGI está comprometido con este cambio de paradigma y trabaja para construir una *cultura de calidad*, de modo que cada decisión sobre procedimientos, capacitación, normatividad y políticas se haga considerando las alternativas que contribuyan a cumplir siempre con sus principios. Esto se basa no sólo en una exigencia técnica, sino también en el entendimiento de que así se contribuye a fortalecer el conocimiento que requerimos para nuestro desarrollo.

<sup>1</sup> Ariza, Francisco Javier, *Calidad en la producción cartográfica*, RA-MA Editorial. Madrid, 2002. Esta definición prácticamente es la misma que maneja, por ejemplo, la Organización Internacional de Normalización (véase norma 19113, *Principios de calidad*, versión preliminar, 2002) y es adoptada en varias fuentes más.

<sup>2</sup> *Op. cit.*, p. 366.

<sup>3</sup> *Ibidem*, p. xxviii.

<sup>4</sup> Reyes, Mario Alberto y Olvera, Jesús, *Infraestructura de datos espaciales*, revista *Este País*, mayo de 2007.

<sup>5</sup> Mas Mayoral, Sebastián, “Normas y estándares para el intercambio de datos geográficos”, en *El campo de las ciencias y las artes*, Servicio de Estudios del banco BBVA, Madrid, 2001.

<sup>6</sup> Reyes, Mario Alberto y Olvera, Jesús, *op. cit.*