





Masó, J. & Pons, X. (2024). Do we need data models beyond raster and vector? The sensor observations case / ¿Necesitamos modelos de datos más allá del ráster y el vectorial? El caso de las observaciones de sensores, *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), 33, 1-5. <http://dx.doi.org/10.21138/GF.862>

---

## DO WE NEED DATA MODELS BEYOND RASTER AND VECTOR? THE SENSOR OBSERVATIONS CASE

### ¿NECESITAMOS MODELOS DE DATOS MÁS ALLÁ DEL RÁSTER Y EL VECTORIAL? EL CASO DE LAS OBSERVACIONES DE SENSORES

<sup>1</sup>JOAN MASÓ  , <sup>2</sup>XAVIER PONS  

<sup>1,2</sup>Grumets Research Group

<sup>1</sup>CREAF. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

<sup>2</sup>Departament de Geografia, Edifici B. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

<sup>1</sup>[joan.maso@uab.cat](mailto:joan.maso@uab.cat) <sup>2</sup>[xavier.pons@uab.cat](mailto:xavier.pons@uab.cat)

Sensor observations (SO) refer to the data collected by sensors that detect and/or measure phenomena in the geographical space. These phenomena can be as different as pedestrians crossing a road, air temperature, concentration of pollutants in a river, movements of nocturnal animals or plant growing. In some cases such as weather radars, the detection or measurement is remote, but the static placement of the sensor combined with a circular scanning creates spatio-temporal datasets quite different in concept from conventional Remote Sensing (RS) systems (even from geostationary sensors). The range and plasticity of SO and their applications make them an increasingly important part of Smart Cities infrastructures (from traffic and noise monitoring and management to intelligent street lighting for energy saving), Precision Agriculture, etc. With the growing presence of the Internet of Things (IoT), conceived as a network of interconnected devices to the Internet, SO collect data and communicate it without human intervention, making possible real time decisions with or without human intervention (with consequent advantages and risks). When the density of observation is enough, it is possible to create a digital twin of the observed “cosmos”, do simulations and predictions and suggest beneficial corrections in the current dynamics and trajectories of the phenomena of interest. As a paradigmatic example, the emergency of low-cost sensors that can measure air pollution at a much lower cost than traditional regulatory monitoring methods is making possible new applications. Indeed, the deployment of air quality monitoring allows better tracking and understanding of pollution patterns and provide additional tools for communities to monitor their local air quality (for instance, by launching citizen science campaigns in synergy with existing regulatory monitoring networks).

One of the first things to learn when approaching a Geographic Information Systems (GIS) for the first time, both as a software, or as a real “system” also including hardware, people, etc, is that there are two basic data models: the raster model and the vector model. The raster data model represents the world as a grid of regularly spaced cells, where each cell holds a value (such as the elevation), or a set of values (such as the radiance sensed by a RS instrument in several spectral bands). The vector model represents the world as a collection of features, where each feature is associated with one or more geometries defined as sequences of coordinates pairs (usually points, lines or polygons) and other properties or values. Both can reflect changes of some properties through time. Neither of these two data models is applicable in all situations, and the two models are considered complementary, the raster being preferable (or essential in practice in cases as RS imagery) to represent variables that change continuously in space (such as elevation, temperature, etc), while the vector is more suitable (or essential in practice in cases as a cadaster dataset) for describing elements that can be individualized by humans

Received: 29/06/2024

*Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)*

[www.geofocus.org](http://www.geofocus.org)

(such as administrative limits, infrastructures, buildings, photo-interpreted forests patches, entire cities as points at very general scales, etc). These two models have been known since the 70s and their consolidation allowed the proliferation of GIS. But neither of the two models is fully appropriate for the data generated by sensors and that is why we have seen the emergence of a specific model to describe this type of data. The reasons behind the need for a new model are:

- \* SO have special metadata requirements to capture the semantics and context of the observations, the sensor type, the measured variable, its units of measurement, etc. The situation can be more complex if we deal with citizen science observation, where we require metadata describing the contributor and what is the license of his/her contributions (that can be potentially different).

- \* Sensor observations can be spatially discrete, with measurements taken at specific locations rather than continuously over a region discarding the raster model.

- \* Sensor observations often have complex temporal characteristics, with measurements not always taken at regular intervals and with varying sampling rates.

- \* Sensor networks can generate massive volumes of data over the same position, something not well supported by a classical vector data model. Instead, SO require efficient storage and powerful querying capabilities according to their nature. In particular, SO are often used in real-time monitoring and decision support systems so data model and storage should provide streaming, data ingestion and rapid access to the most recent observations.

The OGC SensorThings API (STA) part 1 is a standard developed by the Open Geospatial Consortium that combines two elements: an observation-optimized data model and a rich query language. It builds on widely-adopted web protocols and standards such as HTTP Web APIs, JSON, MQTT, and OASIS OData. SensorThings uses HTTP as a main protocol for data access and data gathering, allowing retrieving, creating or deleting of sensor data, but it also adopts MQTT as a data submission and as a subscription mechanism for getting alerts and recent data updates. The data model is based in the OGC Observations and Measurements (that provides a conceptual model) and separate the "thing" and its sensors from the observed feature of interest that is being observed, permitting both in-situ and remote sensing applications. Observations are provided in datastreams that are collections of observations connected to the description of the measured variables and its units of measure. SensorThings adopts the OData as filter and query mechanism and abandons the XML legacy (used in OGC Sensor Observing System) to encode all data in JSON.

In a traditional vector approach, every feature of interest has a property for each measured variable at each time. For example, in meteorology, every weather station is usually conceptualized and implemented as a point feature and in each feature several properties are included in a record, such as the temperature, the relative humidity and the atmospheric pressure, as well as an additional property with the time of the acquisition. The description of these properties, such as the units of measure and the sensor characteristics are stored in metadata. Not many vector models support multiple records associated to the same feature of interest, making the recording of long time series of observations impossible (except if you repeat the feature for each time, or you use solutions with linked tables as Geodatabases or formats like the MiraMon structured vector). In STA, every observation is independent, and it is linked to the feature of interest. Every observed variable is grouped in a different datastream. Being familiar to the traditional vector data model, the STA approach seems unnatural, requiring some experience to formulate the OData queries to get the right data. While the extraction of time series of a single variable is direct, it is surprising how complex is getting the last observations for each feature of interest to create a map, requiring some experience. With time, one learns to appreciate the flexibility of OData to extract the right amount of information needed for each representation... and sometimes requiring to email the server responsible to request re-indexing the database behind the implementation to get responses of complex queries in a reasonable amount of time.

It is expected that the amount of data coming from sensors made available will grow and new applications and services to the communities will emerge. Meanwhile, it will be interesting to see if SO data models and services are adopted by traditional GIS software in the future (for example by implementing STA) and if traditional GIS data models are extended to support the requirements of sensor observations. If that is not the case, sensor data will become the next data silo, remaining disconnected from the rest of the geospatial data.

To learn more about STA and how to formulate OData queries, please visit this excellent tutorial: <https://developers.sensorup.com/docs/>

In this issue we present six interesting works in which Geographic Information Science and Technology are used to evaluate agricultural expansion as an indicator of environmental degradation in Brazil, to understand human pressure in Andalusian coastal areas and to help predict frost risk in the Alacant area. A diagnosis is also presented for the integration of the main Iberian networks of solar radiation stations, as well as an evaluation of the eventual contribution of auxiliary data to the official climatology networks in the Guadalquivir basin. Finally, a multi-criteria evaluation study makes it easier to locate megalithic structures in Aragon.

This editorial was written mainly thanks to the resources provided by the European Union projects CitiObs (HORIZON-CL6-2022-110108642) and AD4GD (HORIZON-CL6-2021-101061001).

---

El término observaciones de sensores (SO, por sus siglas en inglés) hace referencia a los datos recopilados por sensores que detectan y/o miden fenómenos en el espacio geográfico. Estos fenómenos pueden ser tan diferentes como peatones cruzando una calle, temperatura del aire, concentración de contaminantes en un río, movimientos de animales nocturnos o crecimiento de plantas. En algunos casos, como los radares meteorológicos, la detección o medición es remota, pero la ubicación estática del sensor combinada con un escaneo circular crea conjuntos de datos espacio-temporales bastante diferentes en concepto de los sistemas convencionales de Teledetección (incluso de los sensores geoestacionarios). La variedad y plasticidad de los SO y sus aplicaciones los convierten en una parte cada vez más importante de las infraestructuras de las Ciudades Inteligentes (desde la monitorización y gestión del tráfico y el ruido hasta el alumbrado público óptimo para el ahorro energético), la Agricultura de Precisión, etc. Con la creciente presencia de la Internet de las Cosas (IoT, por sus siglas en inglés), concebida como una red de dispositivos interconectados a Internet, es posible recopilar datos y comunicarlos sin intervención humana, posibilitando decisiones en tiempo real con o sin intervención humana (con las consiguientes ventajas y riesgos). Cuando la densidad de observación es suficiente, es posible crear un gemelo digital del “cosmos” observado, realizar simulaciones y predicciones y sugerir correcciones beneficiosas en la dinámica y trayectorias actuales de los fenómenos de interés. Como ejemplo paradigmático podemos señalar las nuevas aplicaciones de medición de la contaminación del aire a un coste mucho menor que el de los métodos tradicionales. En efecto, el despliegue del seguimiento de la calidad del aire mejora el control y comprensión de los patrones de contaminación, a la vez que proporciona herramientas adicionales para que las comunidades sean agentes activos en la vigilancia local (por ejemplo mediante campañas de ciencia ciudadana en sinergia con las redes de seguimiento oficial existentes).

Una de las primeras cosas que hay que aprender al abordar por primera vez un Sistema de Información Geográfica (SIG), ya sea como *software*, o como un “sistema” real que incluye también *hardware*, personas, etc, es que existen dos modelos de datos básicos: el modelo ráster y el modelo vectorial. El modelo de datos ráster representa el mundo como una cuadrícula de celdas espaciadas regularmente, donde cada celda contiene un valor (como la elevación) o un conjunto de valores (como la radiancia detectada por un instrumento de Teledetección en varias bandas espectrales). El modelo vectorial representa el mundo como una colección de entidades, donde cada entidad está asociada con una o más geometrías definidas como secuencias de pares de coordenadas (generalmente puntos, líneas o polígonos) y otras propiedades o valores. Ambos pueden reflejar cambios de algunas propiedades a través del tiempo. Ninguno de estos dos modelos de datos es aplicable en todas las situaciones, y los dos modelos se consideran complementarios, siendo preferible el ráster (o imprescindible a la práctica en casos como las imágenes de Teledetección) para representar variables que cambian continuamente en el espacio (como elevación, temperatura, etc), mientras que el vectorial es más adecuado (o imprescindible a la práctica en casos como un conjunto de datos catastrales) para describir elementos que pueden ser individualizados por los humanos (como límites administrativos, infraestructuras, edificios, áreas de bosques fotointerpretados, ciudades enteras como puntos a escalas muy generales, etc). Estos dos modelos se conocen desde los años 70 y su consolidación permitió la proliferación de los SIG. Pero

ninguno de los dos modelos es del todo apropiado para los datos generados por sensores, y es por ello que hemos visto surgir un modelo específico para describir este tipo de datos. Las razones detrás de la necesidad de un nuevo modelo son:

- \* SO tienen requisitos especiales de metadatos para capturar la semántica y el contexto de las observaciones, el tipo de sensor, la variable medida, sus unidades, etc. La situación puede ser más compleja si nos ocupamos de la observación de ciencia ciudadana, donde requerimos metadatos que describan el contribuyente y cuál es la licencia de sus contribuciones (que pueden ser potencialmente diferentes).

- \* Las observaciones del sensor pueden ser espacialmente discretas, con mediciones tomadas en ubicaciones específicas, en lugar de continuamente sobre una región, descartando el modelo ráster.

- \* Las observaciones de sensores a menudo tienen características temporales complejas, con mediciones que no siempre se toman a intervalos regulares y con velocidades de muestreo variables.

- \* Las redes de sensores pueden generar volúmenes masivos de datos sobre la misma posición, algo que no está bien respaldado por el modelo clásico de datos vectoriales. En cambio, SO requiere un almacenamiento eficiente y poderosas capacidades de consulta de acuerdo con su naturaleza. En particular, los SO se utilizan a menudo en sistemas de soporte de decisiones y monitoreo en tiempo real, por lo que el modelo y el almacenamiento de datos deben proporcionar transmisión, incorporación de datos y acceso rápido a las observaciones más recientes.

La parte 1 de OGC SensorThings API (STA) es un estándar desarrollado por el Open Geospatial Consortium que combina dos elementos: un modelo de datos optimizado para observación y un rico lenguaje de consulta. Se basa en estándares y protocolos *web* ampliamente adoptados como HTTP Web API, JSON, MQTT y OASIS OData. SensorThings utiliza HTTP como protocolo principal para el acceso y la recopilación de datos, lo que permite consultar, crear o eliminar datos de sensores, pero también adopta MQTT para enviar datos y como mecanismo de suscripción para recibir alertas y actualizaciones de datos recientes. El modelo de datos se basa en Observaciones y Mediciones OGC (que proporciona un modelo conceptual) y separa la "cosa" y sus sensores de la característica de interés que se está observando, lo que permite aplicaciones de detección tanto *in situ* como remotas. Las observaciones se proporcionan en flujos de datos que son colecciones de observaciones conectadas a la descripción de las variables medidas y sus unidades de medida. SensorThings adopta OData como mecanismo de filtro y consulta y abandona el legado XML (utilizado en el OGC Sensor Observing System) al apostar por codificar todos los datos en JSON.

En un enfoque vectorial tradicional, cada entidad de interés tiene una propiedad para cada variable medida en cada momento. Por ejemplo, en meteorología, cada estación meteorológica suele conceptualizarse e implementarse como una entidad puntual y, en cada entidad, se incluyen en un registro varias propiedades, como la temperatura, la humedad relativa o la presión atmosférica, así como una propiedad adicional con el momento de la adquisición. La descripción de estas propiedades, como las unidades de medida y las características del sensor, se almacenan en metadatos. No muchos modelos vectoriales soportan múltiples registros asociados a un mismo elemento de interés, lo que imposibilita el registro de largas series temporales de observaciones (excepto si se repite el elemento cada vez, o se utilizan soluciones con tablas enlazadas como Geodatabases o formatos como el vectorial estructurado de MiraMon). En STA, cada observación es independiente y está vinculada a la entidad de interés. Cada variable observada se agrupa en un flujo de datos diferente. Cuando se está familiarizado con el modelo de datos vectoriales tradicional, el enfoque STA parece poco natural y la obtención de los datos deseados requiere cierta experiencia para formular consultas OData correctas. Si bien la extracción de series temporales de una sola variable es directa, sorprende lo complejo que resulta conseguir las últimas observaciones de cada variable de interés para crear un mapa, lo que requiere cierta experiencia. Con el tiempo, uno aprende a apreciar la flexibilidad de OData para extraer la cantidad exacta de información necesaria para cada representación... aunque a veces requiera enviar un correo electrónico al responsable del servidor para solicitar la reindexación de la base de datos existente detrás de la implementación y así obtener respuestas a consultas complejas en un tiempo razonable.

Se espera que la cantidad de datos provenientes de los sensores disponibles crezca y surjan nuevas aplicaciones y servicios para las comunidades. Mientras tanto, será interesante ver si los modelos y servicios de datos SO son adoptados por el *software* SIG tradicional en el futuro (por ejemplo, mediante la implementación de STA) y si los modelos de datos SIG tradicionales se amplían para

soportar los requisitos de las observaciones de sensores. Si ese no es el caso, los datos de los sensores se convertirán en el siguiente silo de datos y permanecerán desconectados del resto de los datos geoespaciales.

Para obtener más información sobre STA y cómo formular consultas OData, visite este excelente tutorial (en inglés): <https://developers.sensorup.com/docs/>

En este número presentamos seis interesantes trabajos en que la Ciencia y la Tecnología de la Información Geográfica se utilizan para evaluar la expansión agropecuaria como indicador de degradación ambiental en Brasil, para conocer la presión humana en zonas litorales andaluzas y para ayudar a la previsión de riesgo de heladas en la zona de Alacant. También se presenta una diagnosis para la integración de las principales redes ibéricas de estaciones de radiación solar, y una evaluación acerca de la eventual contribución de los datos auxiliares a las redes oficiales de climatología en la cuenca del Guadalquivir. Finalmente, un estudio de evaluación multicriterio permite facilitar la localización de estructuras megalíticas en Aragón.

Este editorial fue escrito principalmente gracias a los recursos proporcionados por los proyectos de la Unión Europea CitiObs (HORIZON-CL6-2022-110108642) y AD4GD (HORIZON-CL6-2021-101061001).

## References

Piowar, J. M., LeDrew, E. F., & Dudyca, D. J. (1990). Integration of spatial data in vector and raster formats in a geographic information system environment. *International Journal of Geographical Information Systems*, 4(4), 429-444. <https://doi.org/10.1080/02693799008941557>

Horsburgh, J. S., Aufdenkampe, A. K., Mayorga, E., Lehnert, K. A., Hsu, L., Song, L., Jones, A.S, Damiano, S.G., Tarboton, D.G., Valentine, D., Zaslavsky, I. & Whitenack, T. (2016). Observations Data Model 2: A community information model for spatially discrete Earth observations. *Environmental Modelling & Software*, 79, 55-74. <https://doi.org/10.1016/j.envsoft.2016.01.010>

Liang, S., Khalafbeigi, K., van der Schaaf, H. (2021). OGC SensorThings API Part 1: Sensing Version 1.1, OGC 18-088. Available online: <https://docs.ogc.org/is/18-088/18-088.html>

