





Masó, J., & Pons, X. (2022). After years convincing people about geospatial services, we are going back to files, but this time we call them "Cloud optimized" / Después de años convenciendo a la gente sobre los servicios geoespaciales, volvemos a los archivos, pero esta vez los llamamos "Optimizados para la nube", *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), nº 30, p. 1–4. <http://dx.doi.org/10.21138/GF.813>

AFTER YEARS CONVINCING PEOPLE ABOUT GEOSPATIAL SERVICES, WE ARE GOING BACK TO FILES, BUT THIS TIME WE CALL THEM "CLOUD OPTIMIZED"

DESPUÉS DE AÑOS CONVENCRIENDO A LA GENTE SOBRE LOS SERVICIOS GEOESPACIALES, VOLVEMOS A LOS ARCHIVOS, PERO ESTA VEZ LOS LLAMAMOS "OPTIMIZADOS PARA LA NUBE"

¹JOAN MASÓ  , ²XAVIER PONS  

^{1,2}Grumets Research Group

¹CREAF. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

²Departament de Geografia, Edifici B. Universitat Autònoma de Barcelona. 08193, Bellaterra, Catalonia, Spain

¹joan.maso@uab.cat ²xavier.pons@uab.cat

While probably many of us share the impression that information science & technology advance at full speed and nobody is able to cope with so many new things, sometimes we also feel like if we were trapped in time. Indeed, three decades ago GIS lessons highlighted the importance of file formats and nowadays there is a huge variety of them, as the current 160 GDAL drivers for raster formats, 80 for vector formats, demonstrate. Under the view of data scientists, our attention is on binary formats (usually smaller and faster to read) and how they are structured internally. Most well-designed binary files allow the so-called random access, a term that has nothing to do with randomness, but with direct access: it is used to denote that the file can be accessed to any desired position without reading all its previous content, by opposition to the sequential access of a computer tape, or of a file without any structure allowing this direct access. Having direct access we can use big files, that are conceptually easier to manage (in the same way that the seamless large geographic databases are more appealing than thousands of classic cartographic sheets). The usual key to direct access is a predefined, well designed, structure and some offsets in known positions of that structure that facilitate "jumping" directly to some section of the data without having to read all previous bytes.

With the generalization of the Internet and the use of web services, access a fragment of the information by simply asking for it to a service has been possible since more than a decade; and there was (for the final user or final programmer) no need to think on how to manage files intelligently. In principle, you could ask for a fragment of a pictorial representation of a map or an orthoimage in an OGC Web Map Service, ask for a few entities in a bounding box in an OGC Web Feature Service, or ask for a part of a Digital Elevation Model in a Web Coverage Service. In practice, many early implementations of these services were not well dimensioned, and the performance of such operations was, and sometimes still is, frustrating. The Web Map Tile Service offers a direct and high-performance solution especially for raster GIS data models, but no equivalent tiling services was proposed for general data in *de jure* standards organizations.

The new Internet is dominated by "the Cloud". Instead of setting up services (and their associated data) in their organization infrastructures, suddenly providers need to set up them in

Masó, J., & Pons, X. (2022). After years convincing people about geospatial services, we are going back to files, but this time we call them "Cloud optimized" / Después de años convenciendo a la gente sobre los servicios geoespaciales, volvemos a los archivos, pero esta vez los llamamos "Optimizados para la nube", *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), nº 30, p. 1–4. <http://dx.doi.org/10.21138/GF.813>

cloud infrastructures... and pay for it. Setting up web services into the cloud also needs paying for processing time, what can derive in unexpected costs when the service is more successful than initially planned. The unforeseen solution is to go back to offering files (usually parts of them) directly in the cloud (e.g. in a S3 repository). While everybody was used to the download of entire files, some people turned to a web specification that allows downloading only a fragment of a file (<https://www.rfc-editor.org/rfc/rfc9110#name-range-requests>), because its initial drafts are old enough to be sure it is implemented in the majority of web servers and web browsers. The direct (random) access that was possible in GIS files located in our “physical devices” (local hard disks, resources in a Local Area Network, LAN, etc) is in fact possible in the web. But what is the right format to be used? The community has turned back again to a specification completed many years ago (1992): The TIFF file.

A TIFF can be structured using strips (sequences of rows) or tiles (bi-dimensional small rectangles). In addition, TIFF allows for more than one representation of an image: one at full resolution and one or more reduced resolution images; or offering several bands (regions of the electromagnetic spectrum), as in many Remote Sensing images. Because of this TIFF internal structure, the file can be optimally (directly) accessed to extract subareas. And the community decided to call the combination of internal tiles and reduced resolution images "Cloud Optimized GeoTIFF" (<https://www.cogeo.org/>). Actually, visualizations of COG files in web browsers using the `geotiff.js` library are surprisingly fast (<https://geotiffjs.github.io/>). Now the community is looking for other file formats with similar performance possibilities for other types of data. For instance, Geoparquet, now in 1.0 beta version (<https://github.com/opengeospatial/geoparquet>), wants to become a new format for cloud optimized vector data, structured as tabular data. Nevertheless, perhaps a better solution could come from a file containing vector data structured in vector tiles, able to offer different levels of resolution of the same geometries and complemented by tables of attributes (that could benefit from the parquet format, or from simple table files that allow random access, such as the classic DBF –or its extensions, as <https://www.miramonts.cat/usa/Prod-MiraD.htm>–). More information here: <https://cholmes.medium.com/an-overview-of-cloud-native-vector-c223845638e0>. Seeing the fascination of the community for the COG approach we are going to notice an emergence on several new cloud optimized formats in the next years.

In this issue we present several works of great interest and geographical diversity. In one of them, GIS is used as a multidisciplinary and objective approach in the field of judicial resolutions in a protected urban natural area of Patagonia and reflects on its role. In another, data from social networks in Spain is combined to analyze the movement of people in time and space during the peak of the COVID-19 crisis. A third article presents an extensive review of the implications, for users of land cover maps in regions such as Andalusia, of the fact that they are produced at different scales and geographic areas, as well as with a great disparity in temporal resolutions, legends, and level of thematic detail. The volume is completed with a comparative analysis of road traffic accidents in an area of Ecuador, which will serve to improve awareness and prevention policies, and another analysis, in this case of landscape dynamics in a region of Argentina, carried out through cloud resources, about which we just opened this Editorial article. We hope these works are of interest to you.

We cannot end this Editorial without sharing the good news that the GeoFocus journal has been indexed in Scopus. We would like to express our gratitude to the entire Team of the review (technicians, Associate Editors, Editorial Board and Advisory Board), as well as to the scientists, technicians and public who read and cite us and who, ultimately, are GeoFocus' reason

Masó, J., & Pons, X. (2022). After years convincing people about geospatial services, we are going back to files, but this time we call them "Cloud optimized" / Después de años convenciendo a la gente sobre los servicios geoespaciales, volvemos a los archivos, pero esta vez los llamamos "Optimizados para la nube", *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), nº 30, p. 1–4. <http://dx.doi.org/10.21138/GF.813>

for being. This indexing is added to the many others where the journal has entered, such as the JCR, Emerging Sources Citation Index, or the FECYT quality seal of scientific journals (complete list at <https://www.geofocus.org/index.php/geofocus/indexing>). Congratulations to all.

Si bien probablemente muchos de nosotros compartimos la impresión que la ciencia y la tecnología de la información avanzan a toda velocidad y nadie es capaz de hacer frente a tantas cosas nuevas, a veces también sentimos que estamos atrapados en el tiempo. De hecho, hace tres décadas, las lecciones de SIG destacaban la importancia de los formatos de archivo y hoy en día existe una gran variedad de ellos, como lo demuestran los actuales 160 controladores GDAL para formatos ráster, 80 para formatos vectoriales. Desde el punto de vista de los científicos de datos, nuestra atención se centra en los formatos binarios (generalmente más pequeños y más rápidos de leer) y en cómo se estructuran internamente. La mayoría de los archivos binarios bien diseñados permiten el llamado acceso aleatorio, un término que no tiene nada que ver con la aleatoriedad, sino con el acceso directo: se utiliza para indicar que se puede acceder al archivo en cualquier posición deseada sin leer todo su contenido anterior, por oposición al acceso secuencial de una cinta de ordenador, o de un archivo sin ninguna estructura que permita este acceso directo. Al tener acceso directo podemos utilizar grandes archivos, que conceptualmente son más fáciles de manejar (del mismo modo que las grandes bases de datos geográficas integradas son más atractivas que miles de hojas cartográficas clásicas). La clave habitual para el acceso directo es una estructura predefinida, bien diseñada, y la existencia de algunos valores de desplazamiento en posiciones conocidas de esa estructura que facilitan "saltar" directamente a alguna sección de los datos sin tener que leer todos los bytes anteriores.

Con la generalización de Internet y el uso de servicios web, acceder a un fragmento de la información simplemente solicitándolo a un servicio es posible desde hace más de una década; y no había (ni para el usuario final ni para el programador final) ninguna necesidad de pensar en cómo administrar los archivos de manera inteligente. En principio, podría solicitar un fragmento de una representación pictórica de un mapa o una ortoimagen en un Web Map Service de OGC, solicitar algunas entidades en una envolvente geográfica en un Web Feature Service de OGC, o solicitar una parte de un servicio de un Modelo de Elevaciones en un Web Coverage Service. En la práctica, muchas implementaciones tempranas de estos servicios no estaban bien dimensionadas, y el rendimiento de tales operaciones era, y a veces sigue siendo, frustrante. Web Map Tile Service ofrece una solución directa y de alto rendimiento, especialmente para modelos de datos SIG ráster, pero no se propusieron servicios de mosaico equivalentes para datos generales en las organizaciones de estándares *de jure*.

La nueva Internet está dominada por "la Nube". En lugar de configurar servicios (y sus datos asociados) en las infraestructuras de su organización, de repente los proveedores necesitan configurarlos en infraestructuras en la nube... y pagar por ello. La instalación de servicios web en la nube también requiere pagar por el tiempo de procesamiento, lo que puede derivar en costos inesperados cuando el servicio es más exitoso de lo planeado inicialmente. La solución imprevista es volver a ofrecer archivos (normalmente partes de ellos) directamente en la nube (por ejemplo, en un repositorio S3). Si bien todos estábamos acostumbrados a la descarga de archivos completos, algunas personas recurrieron a una especificación web que permite descargar solamente un fragmento de un archivo (<https://www.rfc-editor.org/rfc/rfc9110#name-range-requests>), ya que sus borradores iniciales son lo suficientemente antiguos como para estar seguros que están implementados en la mayoría de servidores web y navegadores web. El acceso directo

Masó, J., & Pons, X. (2022). After years convincing people about geospatial services, we are going back to files, but this time we call them "Cloud optimized" / Después de años convenciendo a la gente sobre los servicios geoespaciales, volvemos a los archivos, pero esta vez los llamamos "Optimizados para la nube", *GeoFocus, Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Editorial), nº 30, p. 1–4. <http://dx.doi.org/10.21138/GF.813>

(aleatorio) que era posible en archivos SIGI ubicados en nuestros “dispositivos físicos” (discos duros locales, recursos en una Red de Área Local, LAN, etc) es, de hecho, posible en la web. Pero ¿cuál es el formato correcto a utilizar? La comunidad ha vuelto de nuevo a una especificación completada hace muchos años (1992): el archivo TIFF.

Un TIFF se puede estructurar mediante tiras (secuencias de filas) o teselas (pequeños rectángulos bidimensionales). Además, TIFF permite más de una representación de una imagen: una en resolución completa y una o más imágenes en resolución reducida; u ofreciendo varias bandas (regiones del espectro electromagnético), como en muchas imágenes de Teledetección. Debido a esta estructura interna TIFF, se puede acceder de forma óptima (directa) al archivo para extraer subáreas. Y la comunidad decidió llamar a la combinación de mosaicos internos e imágenes de resolución reducida "GeoTIFF optimizado para la nube" (<https://www.cogeo.org/>). En realidad, las visualizaciones de archivos COG en navegadores web que utilizan la biblioteca geotiff.js son sorprendentemente rápidas (<https://geotiffjs.github.io/>). Ahora la comunidad está buscando otros formatos de archivo con posibilidades de rendimiento similares para otros tipos de datos. Por ejemplo, Geoparquet, ahora en versión 1.0 beta (<https://github.com/opengeospatial/geoparquet>), quiere convertirse en un nuevo formato para datos vectoriales optimizados para la nube, estructurados como datos tabulares. No obstante, quizás una mejor solución podría venir de un archivo que contuviera datos vectoriales estructurados en teselas vectoriales, capaz de ofrecer diferentes niveles de resolución de las mismas geometrías y complementado con tablas de atributos (que podrían beneficiarse del formato Parquet, o de archivos de tablas simples que permitan el acceso aleatorio, como el DBF clásico –o sus extensiones, como <https://www.miramont.cat/esp/Prod-MiraD.htm>–). Más información aquí: <https://cholmes.medium.com/an-overview-of-cloud-native-vector-c223845638e0>. Al ver la fascinación de la comunidad por el enfoque COG, pronosticamos la aparición de varios formatos nuevos optimizados para la nube en los próximos años.

En este número presentamos varios trabajos de gran interés y diversidad geográfica. En uno de ellos se utilizan los SIG como abordaje multidisciplinar y objetivo en el ámbito de resoluciones judiciales en un área natural urbana protegida de la Patagonia y se reflexiona sobre su papel. En otro se combinan datos de redes sociales en España para analizar en el tiempo y el espacio el movimiento de las personas durante el pico de la crisis del COVID-19. Un tercer artículo presenta una extensa revisión de las implicaciones que para los usuarios de mapas de cubiertas del suelo en regiones como Andalucía representa el hecho que se produzcan a diferentes escalas y ámbitos geográficos, así como también con una gran disparidad de resoluciones temporales, leyendas y nivel de detalle temático. El volumen se completa con un análisis comparativo de los siniestros de tráfico rodado en un área de Ecuador, el cual servirá para mejorar las políticas de concienciación y prevención, y otro análisis, en este caso de dinámica del paisaje en una región de Argentina, realizado a través de recursos en la nube, sobre los cuales justamente abríamos este artículo de Editorial. Esperamos que estos trabajos sean de su interés.

No podemos terminar este Editorial sin compartir la buena noticia que la revista GeoFocus ha sido indexada en Scopus. Queremos expresar nuestro agradecimiento a todo el Equipo de la revista (técnicos, Editores Asociados, Consejos de Redacción y Asesor), así como a los científicos, técnicos y público en general que nos leen y citan y que, en definitiva, son la razón de ser de GeoFocus. Esta indexación se suma a las otras muchas en las que ha ido entrando la revista, como son JCR, Emerging Sources Citation Index, o el sello de calidad de revistas científicas de la FECYT (lista completa en <https://www.geofocus.org/index.php/geofocus/indexing>). Felicidades a todos.