

The risk of reinventing the wheel in geographic information formats

El riesgo de reinventar la rueda en los formatos de información geográfica

XAVIER PONS  

Grumets research group.

Dep. de Geografia, Edifici B, Universitat Autònoma de Barcelona, Catalunya, España

xavier.pons@uab.cat

Since the early days of Geographic Information Science and Technology in the digital age, such Information has been stored in computer files. The confluence of areas such as remote sensing, CAD, etc, led to extremely varied proposals, partly originated by previous traditions and by the different underlying data models both in terms of space, thematic information, and time, and also because of the need (not always addressed) to achieve efficient access to data. And this has directed to a situation of lack of standardization and, consequently, of lack of interoperability in data formats, only partially mitigated by some efforts in the form of computer libraries (e.g., GDAL/OGR). Indeed, it is evident that a 12-bit per pixel remote sensing image with 12 spectral bands captured regularly in time cannot be effectively or efficiently processed in a text format such as JSON.

After those first proposals there have been many others, more or less ambitious in their flexibility, in their desire to remain simple or to be efficient, to be scalable, to explicitly contain extremely important information in geographic information such as georeferencing, topology, etc. Organizations such as the Open Geospatial Consortium, OGC, have been working in commissions for years to propose *de jure* standards, sometimes adopting *de facto* standards (community standards), sometimes anew. But we also observe or participate in this necessary work from the contradiction that implies a multiplicity of proposals that do not help to stop the dispersion, and that only in some cases achieve success, while in others they are of minor application, being practically ignored independently of their real interest. Why does this double process of multiplicity of proposals, "official" or not, occur? Why does this unreasonable support or forgetfulness occur?

The reasons are multiple, and would require a longer text, but from this editorial it seems appropriate to make a call to not continually reinvent the wheel (e.g., on the pretext that the previous proposal was too complicated). Is it reasonable that now, in the "Tiled Gridded Coverage Version 1.1 extension" for Geopackage, the proposal for raster data in bit depths and attention to the unsigned treatment will not yet include aspects that have been used for years in, for example, MSG-SEVIRI data, so important in meteorology, climate, ecology, etc? Wouldn't it have been better to converge with the proposal of "Cloud Optimized GeoTIFF", COG (which also provides tiles in a much more consolidated format)? Perhaps a moratorium on the definition of new formats would be

Pons, X. (2021): The risk of reinventing the wheel in geographic information formats / El riesgo de reinventar la rueda en los formatos de información geográfica, *GeoFocus (Editorial)*, *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 27, 1-3. <http://dx.doi.org/10.21138/GF.738>

necessary, with two exceptions: the extension of existing formats to cover new needs (*e.g.*, larger files or with support for a larger number of features) and the proposal of new ones when they entail real advances in key aspects (*e.g.*, formats to contain new data models). This would avoid wasting the energy of so many people: those who propose, those who implement, and users who have to discover what the new proposal brings or what disadvantages has.

For various reasons, the geographic information metadata formats have also had a dispersed dynamic, but less so. Let us hope that the sophistications that come to them are to collect real needs currently not covered, and to provide other benefits of which perhaps we will talk another day; in a similar way, this reflection could also be qualified in the case of geoservices if it were not deliberately brief.

In any case, in previous editorials we stated how much we must still work so that sampling and modeling do not require almost guessing processes to know characteristics and thus facilitate more robust scientific methods. Without a doubt the jungle of formats does not help and, with so many, it is everyone's responsibility to think carefully before proposing, before accepting, before distributing and before using.

In this issue of GeoFocus, 27th, our readers will find remote sensing papers that address the study of two simultaneous disturbances in Chilean forests, GIS contributions to the formalization of flood risks, combination of both disciplines for the study of a basin in Brazil, a study on the effect of social programs in different areas of Mexico and an evaluation of the anthropic and natural processes on a field of dunes in Argentina. Likewise, we are pleased to incorporate an article in the framework of teaching in secondary education using GIS to work on concepts and raise awareness about sustainable development.

Desde los primeros tiempos de la Ciencia y la Tecnología de la Información Geográfica en la era digital, dicha Información se ha almacenado en ficheros informáticos. La confluencia de ámbitos como la teledetección, el CAD, etc, conllevó propuestas variadísimas, en parte originadas por las tradiciones previas y por los diferentes modelos de datos subyacentes tanto en lo espacial, como en lo temático, como en lo temporal, y por la necesidad (no siempre atendida) de conseguir un acceso eficiente a los datos. Y esto ha llevado a una situación de falta de estandarización y, consecuentemente, de falta de interoperabilidad en los formatos de datos, sólo parcialmente paliada por algunos esfuerzos en forma de librerías informáticas (*e.g.*, GDAL/OGR). En efecto, es evidente que una imagen de teledetección de 12 bits por píxel y 12 bandas espectrales captada regularmente en el tiempo no puede ser tratada eficazmente ni eficientemente en un formato de texto como un JSON.

Desde aquellas primeras propuestas se han sucedido muchas otras, más o menos ambiciosas en su flexibilidad, en su voluntad de mantenerse simples o de ser eficientes, de ser escalables, de contener explícitamente información importantísima en información geográfica como la georreferenciación, la topología, etc. Organismos como el Open Geospatial Consortium, OGC, llevan años trabajando en comisiones para

Pons, X. (2021): The risk of reinventing the wheel in geographic information formats / El riesgo de reinventar la rueda en los formatos de información geográfica, *GeoFocus (Editorial)*, *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 27, 1-3. <http://dx.doi.org/10.21138/GF.738>

proponer estándares *de iure*, a veces adoptando de estándares *de facto* (*community standards*), a veces de nuevo acuño. Pero también atendemos o participamos en esta necesaria labor desde la contradicción que significa una multiplicidad de propuestas que no ayudan a frenar la dispersión, y que sólo en algunos casos consiguen el éxito, mientras que en otros son de aplicación minoritaria, siendo prácticamente ignoradas con independencia de su interés real. ¿Por qué ocurre este doble proceso de multiplicidad de propuestas, “oficiales” o no, y de soporte u olvido poco racional?

Los motivos son múltiples, y requerirían un texto más extenso, pero desde esta editorial nos parece oportuno hacer una llamada a no reinventar la rueda continuamente (*e.g.*, con el pretexto de que lo anterior era demasiado complicado). ¿Es razonable que ahora se proponga para Geopackage, en la “Tiled Gridded Coverage Version 1.1 extension”, datos ráster en profundidades de bits y atención al tratamiento con y sin signo que no van a recoger todavía aspectos que llevan años distribuyéndose a partir de, por ejemplo, datos MSG-SEVIRI, tan importantes en meteorología, clima, ecología, etc? ¿No hubiera sido mejor converger con la propuesta de “Cloud Optimized GeoTIFF”, COG? (que también proporciona teselas en un formato mucho más consolidado). Quizás fuera necesaria una moratoria en la definición de nuevos formatos, con dos excepciones: la extensión de formatos existentes para cubrir nuevas necesidades (*e.g.*, ficheros mayores o con soporte a un número mayor de entidades) y la propuesta de nuevos cuando conlleven auténticos avances en aspectos clave (*e.g.*, formatos para contener nuevos modelos de datos) para no ir gastando energías de tanta gente: los que proponen, los que implementan y los usuarios que tienen que descubrir qué aporta o qué inconvenientes tiene la nueva propuesta.

Por motivos diversos, los formatos de metadatos de información geográfica han tenido una dinámica también dispersa, pero menos. Esperemos que las sofisticaciones que vayan viniendo para ellos sean para recoger necesidades reales actualmente no cubiertas y para proporcionar otros beneficios de los cuales tal vez hablaremos otro día, como también esta reflexión podría matizarse en el caso de los geoservicios, si no fuera deliberadamente breve.

En cualquier caso, en anteriores editoriales manifestábamos cuánto debemos trabajar todavía para que muestreos y modelizaciones no requieran procesos casi de adivinación para conocer características y así facilitar métodos científicos más robustos. Sin duda la selva de formatos no ayuda y, habiendo tantos, es una responsabilidad de todos pensar bien antes de proponer, antes de aceptar, antes de distribuir y antes de usar.

En este número de GeoFocus, 27, nuestros lectores encontrarán trabajos de teledetección que abordan el estudio de dos perturbaciones simultáneas en bosques de Chile, aportaciones SIG a la formalización de los riesgos de inundación, combinación de ambas disciplinas para el estudio de una cuenca en Brasil, un estudio sobre el efecto de programas sociales en diferentes áreas de México y una evaluación de los procesos antrópicos y naturales sobre un campo de dunas en Argentina. Asimismo, nos complace contar con un artículo en el marco de la docencia en educación secundaria utilizando SIG para trabajar conceptos y concienciar sobre desarrollo sostenible.