

19.Nov.2019

Ref: Incosistencias en datos veredales 2017

Señores DANE

Reciban un cordial saludo,

Primero deseo felicitarlos por las mejoras en el geoportal del DANE y por poner a disposición del público general, datos de calidad con una licencia abierta (<https://geoportal.dane.gov.co/acerca-del-geoportal/licencia-y-condiciones-de-uso/>), datos que son de interés especial para los colombianos, particularmente el mapa veredal 2017.

En este correo deseo reportar algunas inconsistencias que encontré en tal mapa, esperando su retroalimentación, empleando para tal fin el derecho de petición del que gozamos los colombianos de acuerdo al artículo 23 de nuestra Constitución.

Mi nombre es Vladimir Támara Patiño, colombiano que quiere la paz de Dios identificado con C.C: 79'626.331 de Bogotá. Mi dirección de correspondencia es Cr 5 # 33B-02, aunque de ser posible prefiero respuesta por correo. Las inconsistencias que encontré y reporto, se hicieron en tiempo financiado por el proyecto Pasos de Jesús, para el cual puede ir el crédito de las mismas.

1. ANTECEDENTES

Hace un tiempo envié correcciones para el DIVIPOLA 2018 por Twitter (<https://twitter.com/pasosdeJesus/status/1112934788183310336>), no obtuve respuesta pero me alegra encontrar que estas fueron aplicadas en el DIVIPOLA más reciente (el tipo de correcciones que sugerí entonces también las sugiero para veredas como explico a continuación).

2. PROCESO

2.1 Fuente de información

El 18.Nov.2019 desde el sitio del DANE <http://geoportal.dane.gov.co/servicios/descarga-y-metadatos/descarga-nivel-de-referencia-de-veredas/> elegí 2017 y en el enlace que dice SHP descargue una capa con veredas de Colombia.

Del comprimido que se descarga, extraje CRVeredas_2017.dbf que abrí con LibreOffice Calc, guardé como CSV y dejé una copia en https://github.com/pasosdeJesus/sip/blob/datos/CRVeredas_2017.csv

2.2. Análisis de consistencia con DIVIPOLA 2018

Para probar consistencia empleé el DIVIPOLA 2018 anteriormente disponible en: https://geoportal.dane.gov.co/metadatos/historicos/archivos/Listado_2018.xls del cual dejé copia en

https://danzadoresdejesucristo.onlyoffice.com/products/files/doceditor.aspx?fileid=6068570&action=view&doc=WWNOSk1PeXpYVVpOUWhNaVMxUUJsUk9WTzkwV0NUMWpxWjV1VFpCUlpqUT0_ljYwNjg1NzAi0

Hice las pruebas con un sencillo programa que desarrollé en Ruby y del cual dejé fuentes en: <https://gitlab.com/vtamara/watergate/blob/master/analiza-archivo-veredas.rb>

2.3 Resultados de análisis

A manera de resumen, las inconsistencias que encontré fueron:

- En 73 registros el departamento no es correcto
- En 183 registros el municipio no es correcto
- 2930 registros corresponden a veredas con nombre repetido en el mismo municipio (en 70 registros con el mismo código de vereda).
- 2779 registros son veredas con nombre SIN DEFINIR y 3 con nombre SD .
- 93 registros emplean guión, paréntesis o punto de manera errada.

Dejé el detalle de esas incosistencias en la hoja de cálculo disponible en: https://github.com/pasosdeJesus/sip/raw/datos/geografia/problemas_CRVeredas_2017_Dane.ods

Por otra parte encontré que 4034 veredas tienen el mismo nombre que un Centro Poblado en el DIVIPOLA 2018. Eso podría explicarse si todas las veredas pasaron de veredas en 2017 a centros poblados en 2018, pero espero su confirmación. El listado de estas 4034 veredas lo he dejado en: https://github.com/pasosdeJesus/sip/raw/datos/geografia/CRVeredas_2017_Dane_CP_DIVIPOLA_2018.ods

Corregí esa información de manera temporal de la siguiente forma:

- Se corrigieron los 73 departamentos y los 183 municipios que eran diferentes en DIVIPOLA 2018 privilegiando el nombre del DIVIPOLA 2018.
- De las veredas repetidas se detectaron 70 que tenían el mismo código (aunque diferente área), las cuales se eliminaron
- A todas las veredas de nombre repetido se les agregó el código de la vereda al final entre paréntesis (De las veredas repetidas 138 tenían un nombre que no era 'SIN DEFINIR' ni 'SD')
- Se retiraron las veredas que tenían repetido el nombre con el de un centro poblado en el DIVIPOLA 2018.

Deje el archivo resultante en:

https://github.com/pasosdeJesus/sip/blob/datos/geografia/CRVeredas_2017_arreglada_DIVIPOLA_2018_sin_cp_rep.csv

3. SOLICITUDES

- 3.1 ¿Podrían por favor retroalimentarme las inconsistencias que encontré?
- 3.2 ¿Podrían emplear por favor una codificación sin repeticiones? Hay 70 veredas con codificación de vereda repetida
- 3.3 Para mi sería natural que en un mismo municipio no se encuentren 2 veredas con el mismo nombre ni una vereda que se llame igual que un centro poblado. En caso de dos veredas o una vereda y un centro poblado realmente se llamen igual en el mismo municipio ¿podrían por favor agregar algo para distinguirlas (por ejemplo entre paréntesis un consecutivo o el código de la vereda)?
- 3.4 Reitero por este medio una pregunta que hice por Twitter (sin respuesta) sobre nombres de municipios y centros poblados (<https://twitter.com/pasosdeJesus/status/1112935093671268353>) y ahora extendiendo a nombres de veredas ¿paréntesis referencia un nombre antiguo, mientras guion nombres alternativos? ¿Hay alguna publicación donde estén las convenciones para el uso de paréntesis, guión, raya al piso, punto? ¿En alguna publicación incluyen las abreviaturas en los nombres de veredas, e.g KM., Z.U., S. , N.E?
- 3.5 ¿Podrían por favor unificar y cambiar nombre de las veredas SIN DEFINIR y SD para emplear un sólo término (pero preferiblemente distinguiéndolas como indiquen en 3.3)?
- 3.6 Agradecería si pudiera publicar las correcciones que hagan a los datos de las veredas 2017 y me informen para poder emplear los nuevos datos.
- 3.7 ¿Cuando planean publicar la versión 2019 o la siguiente versión del mapa veredal? ¿Con que frecuencia planean publicarlo?
- 3.8 Es realmente difícil hacerle seguimiento a los cambios que hacen en el DIVIPOLA y lo será aún más para el mapa veredal. ¿Podrían por favor emplear formatos abiertos, estándar, planos e independientes de vendedor (como CSV y GeoJSON) como fuente e incluirlos en un control de versiones público? Así quienes empleamos los datos del DANE como referencia podremos replicar los cambios con más facilidad y si aceptarían contribuciones podrían evaluar y aceptar con más facilidad cambios como los que he propuesto.

Reitero mi felicitación por sus esfuerzos por mantener información de calidad, orando para que los puedan continuarlos e incluso incrementarlos.



Vladimir Támara Patiño. C.C. 79'626.331 de Bogotá. WhatsApp: 3165383162