

¿Datos abiertos de la red Cómo vamos? ¿Qué tienen de especial?

Nivel: Básico

Ha sido grande el esfuerzo que desde la Red de ciudades Cómo Vamos (RCCCV) se ha hecho para contar con datos abiertos. Detrás de los archivos y de los datasets terminados, hay un importante trabajo de recopilación, estandarización de los datos para que sea posible su consumo .

Sobre la estandarización de datos

La RCCCV cuenta con una trayectoria de muchos años en 11 ciudades del país, las cuales tienen sus propias características e intereses a la hora de priorizar temáticas adicionales a aquellas que se comparan como Red,. Por esta razón, durante el año 2015 se asumió la tarea de tomar todos los datos de todas las ciudades y homologarlos a para hacerlos comparables en los formatos que requiere ciudatos. . Esto permite hablar de variables cubiertas por la Red.

Sobre la división de los datos y apertura

Al contar con todas las variables subjetivas y objetivas, fue posible pasar a preparar archivos que facilitaran el consumo de la información. Se dividió la información por ciudad, naturalmente; y además se dividió la información por año para facilitar la visualización de las variables disponibles en cada ciudad y como Red.

Los datos tal y como los presentamos buscan ser:

- Abiertos
- Limpios

Sobre los Datos Abiertos

Los datos abiertos, según la Open Knowledge Foundation, deben de cumplir con tres cosas:

- Disponibilidad y acceso: la información debe estar disponible como un todo y a un costo razonable de reproducción, preferiblemente descargandola de internet. Además, la información debe estar disponible en una forma conveniente y modificable.

- Reutilización y redistribución: los datos deben ser provistos bajo términos que permitan reutilizarlos y distribuirlos, e incluso integrarlos con otros conjuntos de datos.
- Participación universal: todos deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que prevendría el uso comercial de los datos; o restricciones de uso para ciertos propósitos (por ejemplo sólo para educación) no son permitidos.

¿Cómo logramos esto? Con formatos que se puedan usar, redistribuir y que no necesiten un programa especial para trabajar con ello. En este caso, usamos CSV, que se parecen mucho a una tabla de Excel, sin la necesidad de tener un programa especial para correrlo. Puedes leer más sobre datos abiertos en el artículo <http://opendatahandbook.org/guide/es/what-is-open-data/>.

Sobre los Datos Limpios

Para que una base de datos sea considerada limpia, se espera que sea clara en:

- Estructura: Se espera que la tabla sea cuadrada, es decir, que los datos estén completos.
- Semántica: Es consistente en referenciar cada valor con su significado. Aclara el significado más profundo de cada elemento mediante un diccionario.
- Valores: Cada columna debe de ser una variable y cada fila una observación

Puedes leer sobre los datos limpios en el artículo <http://vita.had.co.nz/papers/tidy-data.pdf>.