



Recomendaciones uso bases de datos Centro de Estudios

Septiembre de 2014

1. Las bases de datos publicadas en la página del Centro de Estudios se encuentran en formatos zip o rar, que permiten comprimir y adjuntar varios archivos de gran tamaño en un solo archivo. Para descomprimirlos y acceder a la documentación que contienen, debe tener instalado en su computador alguno de los programas descompresores disponibles en internet, tales como Winzip o Winrar.
2. El peso de los archivos comprimidos que aparecen asociados a las bases de datos, no necesariamente refleja el peso real de los documentos que contienen. Existen archivos publicados indicando un peso de “0 kb”, que sí contienen documentación al descargarlos.
3. Respecto al esquema de registro con la información contenida en las bases de datos, el último esquema publicado será el que contiene la mayor cantidad de información y la más actualizada. Por tanto, se recomienda siempre descargar el último archivo y revisar el esquema de registro asociado, independientemente de si trabaja con las bases de datos de años anteriores. Esto porque los esquemas de registro contemplan información acumulativa y los últimos esquemas cuentan con la información disponible para los años anteriores.
4. El nombre de las bases de datos (en formato CSV y/o Excel), contiene, en general, la siguiente estructura: 20140421_directorio_oficial_2013_20130430, donde:
 - a. “20140421” indica la fecha de término de procesamiento de la base de datos o fecha de última actualización o modificación (cuando aplica); en este caso el día 21 de abril de 2014.
 - b. “directorio_oficial” indica el nombre del proceso; en este caso corresponde al directorio oficial de establecimientos.
 - c. “2013” indica el año de la información; en este caso el año 2013, que debe coincidir, además, con la variable “AGNO” al interior de la base de datos.
 - d. “20130430” indica la fecha de corte de información; en este caso “20130430” corresponde a la información capturada hasta el día 30 de abril de 2013.
5. Los formatos predefinidos de las bases de datos pueden ser Excel o CSV. Este último formato permite separar las columnas a través de un “punto y coma” (;).
6. Para la correcta apertura de bases de datos se debe considerar que:
 - a. Excel permite la apertura de hasta 1.048.576 registros (filas), por lo cual se recomienda revisar el esquema de registro antes de su apertura para verificar si el programa podrá abrir la totalidad de los registros.

- b. Para bases de datos con mayor cantidad de registros, se recomienda la utilización de software alternativos, tales como STATA o SPSS.
- c. Si se utiliza el software SPSS se debe considerar:
 - Por defecto, el programa define el ancho de las columnas de variables cadena de acuerdo a la información contenida en las primeras 200 filas. Así, el usuario debe tener en cuenta que puede cortarse la información de cada caso, si utiliza la modalidad predeterminada por este programa.
 - Para la correcta apertura, se debe definir como delimitador para la separación entre variables el carácter “punto y coma” y como calificador de texto “comilla doble”, al mismo tiempo.
- d. Si se utiliza el software STATA se debe considerar:
 - Para la correcta apertura, se debe definir como delimitador para la separación entre variables el carácter “punto y coma”. Por ejemplo, para abrir el archivo *datos.csv* contenido en el directorio Documentos del disco C, se debe escribir lo siguiente en la ventana de comandos de Stata: *insheet using "C:\ Documentos\datos.csv", delimit (";")*El programa al detectar símbolos extraños (no pertenecientes al alfabeto), produce un corte de la información no leyendo todas las filas. Se recomienda siempre chequear con el esquema de registro la correcta lectura de todos los casos de la base de datos.
 - Si el software no permite leer la base completa, se recomienda utilizar el programa STAT TRANSFER para la correcta apertura.