

NOTAS DE CAMPO

EpiData: ¿el heredero natural de EpiInfo 6?

Pedro Arias Bohigas^a / Jens L. Lauritsen^b

^aAsociación SOLES: Software Libre para Epidemiología y Salud Pública, Consejería de Sanidad Comunidad de Madrid, Madrid, España; ^bAsociación EpiData

(EpiData: the natural heir to EpiInfo 6?)

Resumen

EpiData es un programa informático desarrollado por la Asociación EpiData (www.epidata.dk) que, siguiendo la filosofía de EpiInfo 6, pretende ofrecer todas las ventajas de ese programa: sencillez, aplicabilidad, pocos requerimientos del sistema operativo y de los sistemas de comunicación, ampliándolas con un enfoque claro hacia la documentación y la calidad de los datos, y las ventajas que para muchos usuarios tiene el entorno Windows®. El objetivo de esta Nota es presentar a los potenciales usuarios las ventajas y limitaciones de EpiData, que puede ser en breve tiempo el equivalente a lo que EpiInfo 6 fue hace unos pocos años.

Abstract

EpiData is an epidemiological software developed by the EpiData Association (www.epidata.dk). Following the EpiInfo 6 philosophy, EpiData, offers all the advantages of EpiInfo 6: simplicity, applicability, few operation and communication system requirements, widening them with a clear focus on data quality and documentation plus the advantages that, for many users, has the Windows® operating system. The aim of this Note is to introduce to potential users the strengths and limitations of EpiData, a software that can become in a short time the equivalent to what EpiInfo 6 was a few years ago.

EpiData¹ es un programa informático que viene a sustituir, en el entorno Windows, al admirable EpiInfo 6², y ha sido concebido como alternativa al camino seguido por EpiInfo para Windows. EpiData ha procurado mantener la filosofía de EpiInfo 6 incorporando las ventajas que para muchos usuarios tiene el entorno Windows y un decidido enfoque hacia la documentación de los datos. Incluye un gestor de bases de datos y documentación y un módulo de análisis, es de libre distribución, ocupa poco espacio en disco duro y no interfiere con la configuración del ordenador. Hasta el momento de publicar esta Nota, EpiData había sido traducido a los principales idiomas del mundo y el número de descargas superaba las 25.000 (alrededor de 5.000 en la versión en español) en los últimos 2 años. EpiData está siendo utilizado en un amplio rango de estudios y condiciones, algunos ejemplos de los mismos se presentan en la página oficial de la asociación EpiData (www.epidata.dk). El objetivo de esta Nota de Campo es presentar el programa y valorar su utilidad, aplicaciones y limitaciones.

Correspondencia: Dr. Pedro Arias Bohigas.
Asociación Soles. Software Libre para Epidemiología y Salud Pública.
C/ Fuencarral, 67, 4.o izda. 28006 Madrid. España.
Correo electrónico: parias@asoles.org

Recibido: 14 de junio de 2006.

Aceptado: 20 de julio de 2006.

Los módulos de EpiData

Por el momento, EpiData se compone de 2 módulos independientes, uno para la gestión y la documentación de datos (EpiData Entry) y otro para el análisis de datos (EpiData Analysis).

EpiData Entry sigue la filosofía de EpiInfo 6, es decir, trabaja con un fichero de tipo QES donde se definen la estructura de la tabla de datos y la disposición de los campos en la pantalla de grabación; utiliza ficheros de tipo REC para almacenar los datos y permite, opcionalmente, añadir un fichero de tipo CHK con los controles de calidad y codificaciones que desee el usuario. Soporta campos de texto, numérico, fecha y lógicos (como EpiInfo 6), y se han añadido algunos específicos para EpiData (p. ej., encriptados). La mayoría de las órdenes utilizadas en el código *check* de EpiInfo son compatibles con EpiData y se han añadido algunas nuevas, como las enfocadas a realizar copias de seguridad.

El diseño de cuestionarios cuenta con 2 asistentes para la definición de los tipos de campo. El proceso de entrada de datos es tan sencillo como en EpiInfo y se ve favorecido por la posibilidad de usar indistintamente el tabulador, la tecla Intro o el ratón para desplazarse de un campo a otro.

La principal aportación de EpiData en la gestión de datos radica en su enfoque a la documentación, control de calidad y seguridad de los datos³. Desde la po-

sibilidad de incluir etiquetas descriptivas del archivo, las variables y los datos, hasta la documentación de la estructura de la base de datos y del contenido de la misma con un descriptivo sencillo de todos los campos que contenga, pasando por la posibilidad de añadir notas de trabajo mientras se están grabando los datos, EpiData permite dejar constancia documental de casi todas las decisiones, los pasos y el estado de la base de datos en cualquier momento. Además, permite aplicar códigos *check* no sólo durante el proceso de grabación sino también a una base de datos ya existente, de forma que se puedan detectar inconsistencias o errores. Incluye la posibilidad de grabación duplicada (doble entrada) de datos y la posibilidad de comparar el contenido de 2 ficheros. En temas de seguridad, incluye la posibilidad de encriptar datos tanto de variables como de archivos, y se está trabajando en implementar un proceso de trazabilidad que permita dejar constancia de qué cambios se han introducido en la base de datos, cuándo y por qué usuario.

En la tabla 1 se resumen las principales ventajas y limitaciones de este módulo.

EpiData Analysis incluye las herramientas utilizadas más frecuentemente en el trabajo de campo en epidemiología, ya incluidas en EpiInfo 6, y añade otras nuevas. En el apartado de gráficos es evidente la mejora sobre EpiInfo 6, no sólo en su calidad sino en la flexibilidad, la cantidad de opciones incluidas y los tipos de gráficos disponibles. Los resultados son almacenados en formato HTML. En cuanto a su velocidad de proceso de una tabla de datos, es similar a la de EpiInfo 6 y, además, puesto que el entorno de trabajo es Windows, el programa puede gestionar toda la memoria disponible, lo que permite analizar cualquier base de datos sin importar el tamaño. Lee directamente ficheros tipo REC, DBF y CSV. Las órdenes pueden escribirse y ejecutarse directamente desde la línea de órdenes o bien ser seleccionadas desde una barra de herramientas «Flujo de Trabajo» o desde el árbol de órdenes (fig. 1).

Los usuarios de EpiInfo 6 observarán que la mayoría de las órdenes tienen una sintaxis similar a la que estaban acostumbrados, en algunos casos EpiData tiene, para una misma función dos órdenes, una propia y la heredada de EpiInfo 6. Las órdenes pueden guardarse en archivos de programa con la misma extensión que los de EpiInfo 6 (PGM), archivos que pueden ser editados en cualquier procesador de textos tipo Bloc de Notas de Windows.

Muchos de los resultados intermedios son almacenados en variables internas del programa, que un usuario un poco avezado podrá utilizar para realizar otros cálculos de forma relativamente sencilla. Como en EpiData Entry, EpiData Analysis aporta un valor añadido en relación con la calidad. Cuando un archivo de datos es leído, EpiData realiza un control de calidad en los datos, detectando y advirtiendo al usuario de posibles

Tabla 1. Ventajas y limitaciones de EpiData Entry

Ventajas

Enfocado a la documentación de datos: etiquetas, notas, etc.
Control de calidad: doble entrada, control de consistencia interna, fácil programación de valores legales y saltos condicionales
Seguridad de datos y confidencialidad: *backup* automático, *backup* encriptado de datos (estándar DES), encriptado de archivos
Ampliable por el usuario: funciones de control definidas por el usuario, llamada a programas externos
Compatibilidad: *open standard*, exporta con etiquetas a Stata, SPSS y SAS; sólo datos a CSV y DBF; casi 100% compatible con EpiInfo 6
No interferencias con el sistema operativo: no depende del registro; puede funcionar desde una llave USB
Entornos: IBM compatibles, Mac y UNIX

Limitaciones

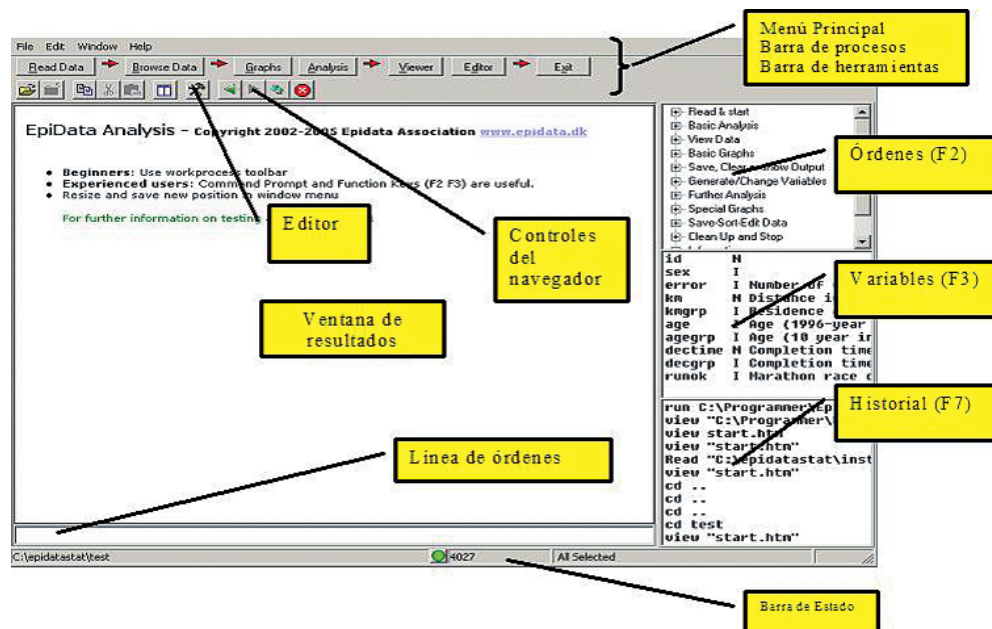
Monousuario: sólo un usuario puede acceder a un archivo de datos cada vez.
El programa sí puede emplearse por varios usuarios simultáneamente
Diseño de pantallas: orientado a líneas, sólo puede definirse una fuente y color por cada archivo
Tipos de campos: no hay campos multitexto (memo)
Control de edición: no es posible realizar un seguimiento de quién y cuándo se ha editado un registro (*audit trail*)
Tamaño de archivos: si el archivo QES tiene más de 64 Kb, deberá ser editado fuera del programa
Tamaño de cadenas de texto: 80 etiquetas de variables, 30 etiquetas de valores, 80 caracteres de campos de texto

errores, como fechas imposibles o muy extremas; además, incorpora automáticamente las etiquetas para los datos definidas por el usuario. La versión disponible en este momento no incluye, sin embargo, algunas de las herramientas de análisis habituales en el mundo de la epidemiología. Por ejemplo, EpiData no permite por el momento realizar regresión logística o análisis de supervivencia.

¿Qué aporta EpiData?

EpiData ha sido diseñado con la idea de servir como herramienta en el trabajo de campo en epidemiología, no pretende sustituir ni a los programas de gestión de bases de datos profesionales ni a los paquetes estadísticos comerciales. Su fortaleza está en la aplicación directa al trabajo rutinario básico, así como su enfoque al control de calidad y documentación de los datos. Incluye en un solo programa todas las herramientas básicas del trabajo de campo en epidemiología. También destaca su sencillez de uso, que lo convierte en una herramienta ideal para la docencia⁴. Puesto que los ficheros de datos son ficheros de texto plano, su tamaño es bastante pequeño, lo que puede suponer una importante ventaja en países donde las comunicaciones son de mala calidad, a la hora de transmitir datos electrónicamente. Su principal debilidad radica en la de-

Figura 1. Ventana principal del módulo de análisis de datos de EpiData. Versión en inglés.



pendencia de un grupo de usuarios organizados en torno a una asociación y, por tanto, con fuentes de financiación limitadas e inconstantes. En este sentido, la asociación EpiData busca el apoyo de las organizaciones de salud pública, ya sean dedicadas a la vigilancia y a la epidemiología de campo o a la docencia, para que colaboren económicamente en el desarrollo y la mejora del programa.

Para los usuarios que conocieron EpiInfo 6 se puede decir que ya tienen un digno y legítimo heredero en entorno Windows. Para los que no han conocido EpiInfo 6 y necesitan iniciarse en el mundo del diseño, gestión y documentación de bases de datos y en el análisis de los mismos desde la epidemiología y la salud pública, EpiData puede ser su herramienta. Es sencillo de utilizar, es de libre distribución y hace casi todo lo que un epidemiólogo necesita.

En la actualidad el módulo de EpiData Entry está traducido al castellano y ya están disponibles algunos materiales de autoformación. El módulo de análisis todavía no había sido traducido en el momento de escribir esta nota, pero se había comenzado la traducción de toda la documentación del programa incluyendo las ayudas.

¿Cómo obtener EpiData?

EpiData puede descargarse de la página oficial de la Asociación EpiData (www.epidata.dk), que desde hace poco se encuentra traducida al español (aparta-

do Download o Descargas). Allí encontrará los enlaces necesarios para obtener los archivos ejecutables de instalación, uno para EpiData Entry y otro para EpiData Analysis. Una vez descargados estos archivos, haciendo doble clic sobre cada uno de ellos y siguiendo las instrucciones el programa quedará instalado. Puede encontrar material de autoaprendizaje y ejemplos sobre el uso del programa en la misma página (apartados de «Documentación» y «Ejemplos»). Si necesita ayuda específica, consulte el apartado Ayuda, donde encontrará distintas estrategias para obtener respuesta a sus dudas.

Bibliografía

1. Lauritsen JM, editor. EpiData Data Entry, Data Management and basic Statistical Analysis System. Odense Denmark, EpiData Association, 2000-2005. Disponible en: <http://www.epidata.dk>
2. Dean AG, Dean JA, Coulombier D, Brendel KA, Smith DC, Burton AH, et al. Epi Info, Version 6: a word processing, database, and statistics program for public health on IBM-compatible microcomputers. Atlanta:Centers for Disease Control and Prevention; 1996.
3. Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med. 2005;2:267. Disponible en: <http://medicine.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pmed.0020267>
4. International Union Against Tuberculosis and Lung Disease. Curso de EpiData [consultado 12/9/2005] 12 Sep 2005]. Disponible en: www.tbrieder.org/epidata/epidata_course/epidata_course.pdf