

Estructura del Servidor Interactivo de Información Epidemiológica.

S. Rodríguez Fernández^{(1),(2)}, R. Fernández-Cuenca Gómez^{(4),(2)}, G. López-Abente Ortega^{(3),(2)} y R. Cano Portero^{(1),(2)}

(1) Área de Datos Epidemiológicos, Centro Nacional de Epidemiología, Instituto de Salud Carlos III, C/ Sinesio Delgado, 6, 28029 Madrid, srofer@isciii.es, rcano@isciii.es

(2) CIBER Epidemiología y Salud Pública (CIBERESP)

(3) Área de Epidemiología Ambiental y Cáncer, Centro Nacional de Epidemiología, Instituto de Salud Carlos III, C/ Sinesio Delgado, 6, 28029 Madrid, glabente@isciii.es

(4) Área de Análisis Epidemiológico y Situación de Salud, Centro Nacional de Epidemiología, Instituto de Salud Carlos III, C/ Sinesio Delgado, 6, 28029 Madrid, rfcuenca@isciii.es

RESUMEN

Los sistemas de software SIG Open Source implementados en servidores Web se encuentran en una etapa de madurez, de forma que son en muchos aspectos iguales o superiores a sus homónimos comerciales; su sofisticación, solidez, estabilidad y manejabilidad facilitan la creación de sistemas de información geográfica accesibles mediante Internet, especializados en diferentes materias, entre ellas la investigación médica. La vigilancia de la situación en salud sigue siendo la piedra angular de la detección de amenazas para la población, que requieren medidas de salud pública. La distribución de este tipo de información especializada es muy útil tanto para apoyar posteriores estudios técnicos, como para cubrir las necesidades de información de los potenciales usuarios. El Centro Nacional de Epidemiología ha desarrollado el Servidor Interactivo de Información Epidemiológica (SII), un SIG basado en WEB con tecnología cliente-servidor. De esta forma, el usuario accede a través de Internet al servidor Apache, soportado con Debian y programado en PHP, lo que permite al usuario interactuar dinámicamente mediante el uso de sentencias MySQL con la base de datos que contiene información de salud referenciada espacialmente. El usuario tiene la posibilidad de escoger el elemento que desea visualizar y la forma de visualizarlo, ya sea en mapas proporcionados por MapServer o en diferentes tipos de gráficas que facilita OWTChart, acompañados por sus tablas de datos. Con el paquete estadístico R conseguimos la suavización de las tendencias temporales.

Se ha conseguido desarrollar una plataforma interactiva, de fácil acceso y sencillo manejo, cuyo objetivo no es que los usuarios obtengan un mapa o una gráfica, sino la respuesta a una pregunta formulada. Las principales aplicaciones del SIIE radican en el ámbito de la investigación científica, la educación sanitaria, como sistema generador de hipótesis y como plataforma de información al público en general.

Palabras clave: *SIG libre, servidor, MapServer, Debian, MySQL, OWTChart, código abierto, cliente/servidor, información epidemiológica, salud pública.*

ABSTRACT

The GIS Open Source Software systems implemented in Webservers are at a stage of maturity, sophistication, robustness, stability and manageability, this makes it easy to create geographic information systems to access through Internet, specializing in different subjects, including medical research. Monitoring the health situation remains the cornerstone of the detection of threats to the population, which require public health measures. The distribution of specialized information is useful both to support further technical studies, to meet the information needs of potential users.

The National Epidemiology Center has developed the Epidemiological Information Interactive Server (SIIE), a GIS Web-based with client/server environment technology. This way, the user access through Internet to the support for Apache server on Debian with PHP programs, which allows the user to interact with dynamic content to MySQL databases through the use of MySQL sentences which containing spatially referenced of health information. The user can choose the item that they want to display and how to display it, in either maps provided by MapServer or in different types of charts that facilitates OWTChart, all that accompanied by data tables. With the statistical package R got the smoothing of time trends. It has developed an interactive, accessible and easy operation, whose goal isn't to give to the users a map or a chart, but it gives the answer to a question made. The main applications of SIIE are in the scientific research field, health education, as a generating hypotheses system and as a platform for take out information to general people.

Key words: *openGIS, server, MapServer Debian, MySQL, OWTChart, open source code, client/server, epidemiologic information, public health.*

INTRODUCCIÓN

La identificación y evaluación de riesgos relacionados con la aparición de enfermedades en la comunidad es un objetivo prioritario en la investigación de salud pública y, más concretamente, en el Centro Nacional de Epidemiología (CNE). El análisis espacial de la distribución de las enfermedades mediante la generación de información geográfica es uno de los elementos más relevantes en la identificación de problemas de salud pública.

El desarrollo del conocimiento científico está íntimamente ligado a la obtención de datos fiables y de calidad. Habitualmente se almacena gran cantidad de información de la cual se necesita extraer aquella que es útil. La creación de una herramienta que aúne la accesibilidad a las fuentes de información y la metodología de análisis potencia la capacidad de los grupos de investigación.

El CNE, perteneciente al Instituto de Salud Carlos III, puso en marcha en 2004 un proyecto para el desarrollo de un Servidor Interactivo de Información Epidemiológica (SIIE) [1], un Sistema de Información Geográfica (SIG) basado en WEB con tecnología cliente-servidor que permite a los investigadores en salud pública y al resto de usuarios descargar información espacio-temporal personalizada sobre tendencias, así como la distribución geográfica de la mortalidad causadas por diversas patologías [31]. Inicialmente, dicho proyecto se ideó para cubrir las necesidades del área de epidemiología ambiental y cáncer pero en la actualidad el SIIE se ha extendido hasta abarcar diversas áreas del CNE, como son; el área de análisis epidemiológico y situación de salud, o el área de vigilancia de epidemiológica y las enfermedades infectocontagiosas.

El objetivo para el que se ha creado el SIIE es facilitar información epidemiológica de calidad, múltiple y variada, de una manera rápida, eficaz y adecuada a las necesidades de cualquier usuario, que sirva de base para posteriores investigaciones científicas, para informes técnicos o incluso como recurso docente, proporcionando un valioso apoyo al ámbito de la salud pública.

El proyecto adopta como modelo el desarrollado por el departamento de vigilancia de enfermedades on-line de la Agencia de Salud Pública de Canadá [3], tanto en su ingenioso diseño como en las herramientas informáticas que se adoptaron en su constitución. El servidor canadiense muestra datos sobre mortalidad de cáncer, enfermedades cardiovasculares, enfermedades crónicas, enfermedades de declaración obligatoria y vigilancia de lesiones. La Agencia de Salud Pública de Canadá, que se engloba dentro del sistema de salud pública canadiense, trabaja con médicos, técnicos de laboratorio, autoridades de salud pública y organizaciones profesionales para mejorar los equipos médicos de primera línea [4].

El objetivo de este trabajo es presentar la estructura interna del SIIE, su desarrollo con software libre y resaltar su importancia en el ámbito de la investigación en salud pública en España y como fuente fiable de información para cualquier usuario interesado en este campo.

MATERIAL Y MÉTODOS

Se ha constituido un sistema con una estructura lógica completamente integrada, dotada de funciones que trabajan sincronizada y conjuntamente, de forma que sus

aplicaciones y unidades funcionales están perfectamente comunicadas e interactúan entre sí. Se trata de un sistema de información dinámico, con aplicaciones hechas a medida y que han sido desarrolladas con la estrecha colaboración de epidemiólogos e informáticos.

Tanto las bases de datos como las aplicaciones del proyecto se han implementado utilizando software de código abierto. Esta opción ha sido elegida por las ventajas que aporta software libre, junto con su potencia, fiabilidad, facilidad tanto de manejo, aprendizaje y adaptación, así como por la existencia de productos ya desarrollados y probados en el ámbito sanitario.

Recursos de entorno

A continuación se exponen los componentes hardware, así como las herramientas Software, utilizados para la instalación del servidor, que unidas proporcionan la plataforma de soporte base de la aplicación desarrollada.

Herramientas de desarrollo

El SIIE es un Servidor **LAMP** que integra las tecnologías **Linux**, **Apache**, **MySQL** y **PHP**, que constituye un paquete de software libre robusto, rápido de implementar, sencillo de mantener y completamente gratis. El conjunto de subsistemas de software necesarios para configurar ésta solución web y conseguir un servidor dinámico de esfuerzo reducido son los siguientes:

- **Debian 4.0** [5], actúa como núcleo. Se trata de un sistema operativo tipo Linux/GNU que, habiendo heredado la robustez y confiabilidad de Unix, posee las ventajas de ser portable, multitarea y multiusuario, lo que lo hace muy potente. Funciona en la mayoría de las plataformas hardware con requerimientos mínimos y tiene una gran capacidad para soportar cargas y aplicaciones con altísimos niveles de exigencia.
- Servidor http **Apache** 2.2.3 [6], posibilita los contenidos dinámicos, a la vez que aporta al sistema estabilidad y velocidad de respuesta. Se trata de un servidor web de código abierto que tiene las ventajas de ser modular, multi-plataforma, extensible, además de tener un soporte para CGI (*Common Gateway Interface*), que posibilita características extendidas como personalización de las variables de entorno y soporte de reparación de errores o *debugging*.
- Motor de bases de datos **MySQL** 5.0.32 [7], que interactúa eficazmente con el sistema para la obtención de datos. Se trata de un Sistema de Gestión de Bases de Datos (SGBD) relacional, multihilo y multiusuario. Brinda grandes posibilidades al ser muy rápido, eficiente, estable y confiable.
- Lenguaje de programación **PHP** 5.2.0 (Hypertext Preprocessor) [8] normal y devel, utilizados del lado del servidor, a través de páginas web. Provee de interactividad y dinamismo al SIIE.
- Generador de gráficos **OWTChart** 1.0.4 [9], que produce los diferentes tipos de gráficas en forma de imágenes GIF a partir de un conjunto de parámetros de entrada. Es utilizado como un CGI en el entorno del servidor web, devolviendo la imagen construida a partir de los parámetros encontrados en el *QUERY_STRING* como parte de la URL. Se basa en la librería GDChart, que a su vez usa la librería Boutell.com's GD (GifDraw) [10]. El motor de gráficas OWTChart ha sido desarrollado parcialmente en virtud de un contrato con Ministerio de Salud de Canadá como parte de su sistema *Disease Surveillance On-Line project*.
- **R** [11], es un software de análisis estadístico de código abierto basado en el

lenguaje S, desarrollado a principio de los años 90. El paquete estadístico R es uno de los más flexibles, potentes y profesionales que existen actualmente para realizar tareas estadísticas de todo tipo, desde las más elementales hasta las más avanzadas. Su aplicación está especialmente recomendada en el campo de las ciencias biomédicas. Cuenta, además, con la ventaja de ser gratuito y de descarga e instalación sencillas.

En el sistema R es utilizado en el "back end" y enlazado mediante el paquete **CGIwithR** [12], que permite utilizar este programa como un lenguaje de *script* CGI, de forma que aplica la función `lowess()` sobre los resultados proporcionados por la consulta, y realizar el suavizado de las tasas, consiguiendo mejorar las salidas de las tendencias temporales.

- **MapServer 4.0.1** [13] genera los mapas interactivos provinciales, municipales y por comunidades autónomas. Se trata de un entorno de desarrollo en código abierto (*Open Source Initiative*) para la creación de aplicaciones SIG en Internet/Intranet, con el fin de visualizar, consultar y analizar información geográfica a través de la red mediante la tecnología *Internet Map Server* (IMS). Soporta los formatos vectoriales ESRI *shapefiles*, PostGIS [31], ESRI ArcSDE [31], GML [31] y otros muchos vía OGR [31], los formatos raster JPG[31], PNG [31], GIF, TIFF/GeoTIFF [31] [31], EPPL7 [31] y otros vía GDAL, así como las fuentes *TrueType*.

Mapserver está ampliamente documentado en la red; para una rápida introducción a los principales conceptos que subyacen detrás de MapServer se puede acceder a MapServer Tutorial [19]. Entre la gran cantidad de información accesible en internet, está MapServer Recipes [15], que contiene ejemplos sencillos de algunas tareas básicas con MapServer.

Funcionamiento de MapServer

Mapserver funciona como una aplicación CGI, instalada en el servidor http, estableciendo la comunicación entre el servidor Web y el programa, de tal modo que este último puede interactuar con Internet.

Los mapas son generados mediante la superposición y fusión de las capas de información solicitada según las opciones seleccionadas en el formulario. Las librerías instaladas le dan la capacidad de realizar diversas acciones sobre el mismo, como un *zoom* sobre una zona.

El CGI de Mapserver utiliza los siguientes recursos:

El principal archivo en la configuración de MapServer es el archivo **Mapfile**[16], que define los datos a ser usados en la aplicación, muestra y consulta de parámetros. **Incluye una serie de parámetros que definen las capas disponibles para el mapa interactivo, el estilo con que se mostrarán esas capas, su simbología, el formato en que se generarán las imágenes, el sistema de referencia, tamaño de las imágenes, entre otros.** También contiene información acerca de cómo se debe dibujar el mapa, la leyenda y el resultado de realizar una consulta, controlando la forma en que las salidas de mapas y las leyendas de Mapserver se deben presentar en la página html. Habitualmente lleva la extensión `.map`.

La extensión **PHP/Mapscript** [17] para PHP, que permite acceder a la API de MapServer a través del *framework* de clases.

El **Template File** controla la aplicación de Mapserver en la ventana del browser,

permitiendo definir la posición de presentación del mapa, la leyenda y determina que vías son disponibles para que el usuario interactúe con Mapserver (browse, query, zoom, etc.).

El **archivo de inicialización** se usa para enviar una consulta inicial al servidor http, que retorna un resultado del servidor de mapas. Permite pasar una variedad de parámetros a la aplicación como ocultos o hidden.

El CGIMapServer necesita para su funcionamiento un **conjunto de Datos SIG**. Nuestra base de datos espacial está compuesta por archivos con formato **shapefile**. El **formato ESRI Shapefile (SHP)** [19] es un **formato vectorial de almacenamiento digital de datos espaciales, de código abierto, donde se guarda la localización de los elementos geográficos y los atributos asociados a ellos. Desarrollado por la compañía ESRI**, compañía que desarrolla y comercializa software para Sistemas de Información Geográfica como ArcInfo o ArcGIS. Originalmente se creó para la utilización con su producto ArcView GIS, pero **actualmente se ha convertido en formato estándar de facto** para el intercambio de información geográfica entre SIGs por la importancia que los productos ESRI tienen en el mercado SIG y por estar muy bien documentado. [19]

Un shapefile es generado por varios archivos relacionados entre sí, con el mismo nombre y las extensiones siguientes:

- * **.shp** - es el archivo que almacena las entidades geométricas de los objetos.
- * **.shx** - es el archivo que almacena el índice de las entidades geométricas.
- * **.dbf** - es la base de datos o archivo que almacena la información de los atributos de los objetos.

Para que todos los componentes interactúen correctamente entre sí y para ofrecer salidas optimizadas, se han instalado varias librerías. De entre las diferentes librerías para el manejo e intercambio de gráficos, cabe destacar la librería **GDAL** [20], que es una biblioteca de acceso a datos *raster* que soporta multitud de formatos y es multiplataforma.

La siguiente figura muestra, de manera gráfica y visual, los distintos componentes del sistema y cómo interactúan entre ellos.

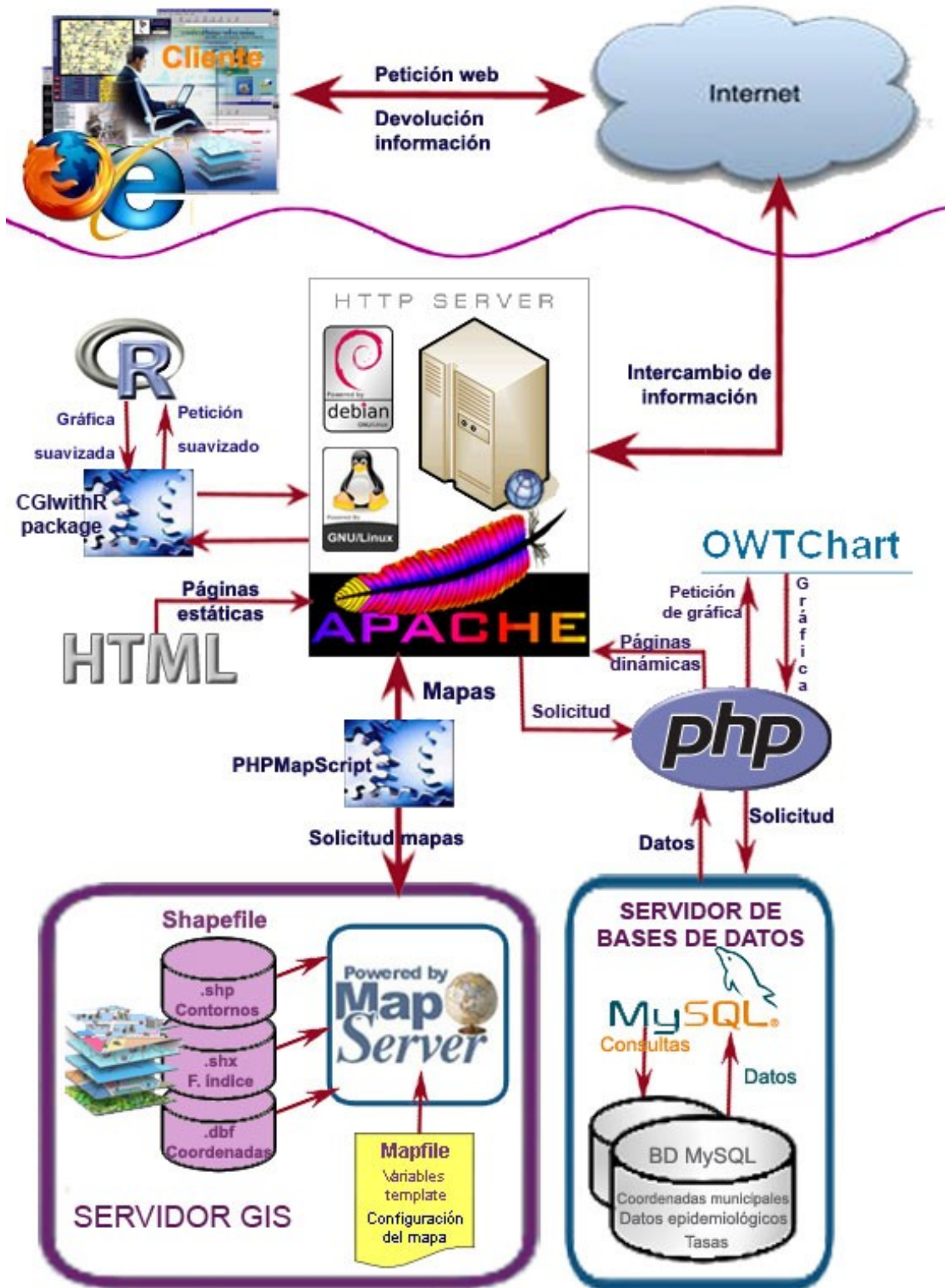


Figura 1: Arquitectura del sistema.

Componentes hardware

Se trata de un sistema que tiene una alta escalabilidad, por lo que puede operar en equipos de prestaciones modestas y, a partir de ahí, es posible migrarlo a instalaciones superiores, sin pérdida de información ni de infraestructura.

Ninguno de los componentes del sistema tienen altos requerimientos de hardware, exceptuando el paquete estadístico R que, para trabajar con grandes masas de datos, necesita que la capacidad de memoria sea como mínimo de 1 GB RAM. Para los demás, es suficiente con satisfacer los requerimientos recomendados para instalar el sistema operativo que se vaya a emplear, dependiendo de la máquina.

El desarrollo del sistema se ha implementado con un servidor Pentium IV de 1.7 MHz y 1 GB de RAM, con Debian 4.0. Es importante destacar que hubiera sido posible dicho desarrollo con menos requisitos, como se puede comprobar en los manuales de instalación de los paquetes mencionados. No obstante, teniendo en cuenta que el sistema en producción proporciona servicio a un número relativamente grande de usuarios concurrentes y, que tanto las gráficas como los mapas se crean en tiempo real, los requerimientos mencionados anteriormente contribuirán a poder trabajar de manera más cómoda y permitirán su expansión a las diversas áreas de investigación integradas en el CNE.

Unidades funcionales

El SIIE consta actualmente de dos aplicaciones totalmente operativas, Ariadna y Raziel, y una tercera aplicación, desarrollada aunque todavía no operativa, llamada Ondina.

El sistema permite en todas ellas, además, la descarga de datos en tablas con formato *.txt, ó *.xls y la extracción del material gráfico en formatos *.gif, *.png ó *.jpg, para su aplicación en estudios posteriores.

Aplicación Ariadna

Esta aplicación genera mapas, gráficas y tablas de datos que muestran la mortalidad por cáncer y otras causas desde el año 1975 hasta 2005, para 60 causas de muerte, por sexo y provincia. Para la creación de la base de datos que es consultada por la aplicación, se utilizó como fuente inicial de información los ficheros de mortalidad del Instituto Nacional de Estadística (INE) [21].

Aplicación Raziel

Esta aplicación genera mapas, gráficas y tablas de datos que muestran las diferencias en la mortalidad para 127 causas de muerte, por sexo, grupo de edad y comunidad autónoma, desde el año 1980 hasta 2005. Para la creación de la base de datos, también se utilizó como fuente inicial de información los ficheros de mortalidad del INE [21].

Aplicación Ondina

Esta aplicación proporciona información sobre el número de casos incidentes, así como sus tasas brutas de incidencia por 100.000 habitantes de enfermedades objeto de vigilancia notificados al sistema de vigilancia epidemiológica. Los datos están disponibles por provincias, comunidad autónoma y semana epidemiológica.



Figura2: interfaz del web site

Calculo de las tasas

Para el cálculo de las tasas se han utilizado los denominadores poblacionales correspondientes a la serie de años completa disponible en el Centro Nacional de Epidemiología. Las referencias poblacionales utilizadas para ello son:

1971-1980: Evolución de la población de España entre los censos de 1970 y 1981, Instituto Nacional Estadística.

1981-1990: Evolución de la población de España entre los censos de 1981 y 1991, calculada por el Instituto Nacional Estadística.

1991-2001: Población calculada en el Centro Nacional de Epidemiología por método de interpolación geométrica, tomando como años pivotaes 1990 y 2002 de las estimaciones del Instituto Nacional Estadística.

2002-2015: Proyecciones de población, calculadas por el Instituto Nacional Estadística, a partir del Censo de 2001.

CONCLUSIONES

La existencia de las especificaciones abiertas de diseño e implementación tales como las definidas por el OGC, ha propiciado que algunos proyectos de desarrollos de software libre hayan podido alcanzar un buen nivel de madurez y estén siendo utilizados en múltiples organizaciones tanto públicas como privadas.

Con la puesta en marcha del SIIE en el CNE-ISCIII en 2004, y sus posteriores ampliaciones, se ha conseguido desarrollar una plataforma interactiva, de fácil acceso y manejo, cuyo objetivo no es que los usuarios obtengan un mapa o una gráfica, sino la respuesta a una pregunta formulada. Su interfaz sencilla facilita el acceso a la información básica en diferentes formatos de gráficos y tablas. Las robustas herramientas, de código abierto y gran estabilidad, son muy adecuadas para entornos profesionales públicos y de investigación, reduciendo el tiempo de implementación y su coste.

La puesta en marcha de un servidor interactivo que proporcione información epidemiológica es un proyecto innovador en el campo de la investigación epidemiológica por las técnicas de análisis, recuperación y representación de información utilizadas. Su base de datos almacena información en una auténtica base de conocimiento, relacionando los datos y estableciendo relaciones entre los mismos. Supone un punto de encuentro de investigadores en el campo de la salud pública.

Desde su puesta en marcha el servidor se ha validado como sistema eficiente por la gran difusión de sus contenidos. La gran innovación del sistema consiste en poder “abrir” el conocimiento de la de la información del campo de la salud pública a todos los usuarios interesados, que no tienen por qué ser necesariamente especialistas, compartiendo informaciones útil.

Sus principales aplicaciones radican en el ámbito de la investigación, educación sanitaria, como sistema generador de hipótesis y como plataforma de información al público general. El acceso por Internet garantiza su difusión.

AGRADECIMIENTOS

Este proyecto está financiado en parte el **Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP)** y el Fondo de Investigación Sanitaria (FIS), Expediente C03-09.

REFERENCIAS

http://www.isciii.es/htdocs/centros/epidemiologia/epi_servidores.jsp
<http://dsol-smed.hc-sc.gc.ca/dsol-smed/>
<http://www.debian.org/>
<http://www.apache.org/>
<http://www.mysql.com/>
<http://www.php.net/>
<http://www.maptools.org/owtchart/>
<http://www.boutell.com/>
<http://www.r-project.org/>
<http://biometry.gis.umn.edu/tutorial/>
<http://www.mobilegeographics.com/mapserver/>

<http://www.mapserver.org/mapfile/>
<http://www.mapserver.org/documentation.html>
<http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>
<http://www.esri-es.com/>
<http://www.gdal.org/>
<http://www.ine.es>
<http://www.opensource.org/>
<http://postgis.refractor.net/>
<http://www.esri.com/software/arcgis/arcscde/>
<http://www.opengis.net/gml/> GML <http://www.gdal.org/ogr>
<http://www.jpeg.org/>
<http://www.libpng.org/pub/png/>
<http://partners.adobe.com/public/developer/tiff/index.html> TIFF/GeoTIFF
<http://www.awaresystems.be/imaging/tiff.html>
<http://www.lmic.state.mn.us/EPPL7/EPPL7/index.html> EPPL7
