

## EL USO DE INTERNET CON SOFTWARE LIBRE Y FUENTES ESPACIALES ABIERTAS PARA COLABORAR EN LA TOMA DE DECISIONES

G. BRENT HALL<sup>1</sup>, JUAN PABLO ALPERIN<sup>2</sup> y SANTIAGO KERRIGAN LEÓN<sup>3</sup>

<sup>1</sup> School of Surveying. University of Otago. Dunedin, New Zealand

<sup>2</sup> Facultad de Geografía. Universidad de Waterloo. Ontario, Canadá N2L 3G1

<sup>3</sup> Centro Alternativa San Martín de Porres. Lima, Perú

<sup>1</sup>[brent.hall@surveying.otago.ac.nz](mailto:brent.hall@surveying.otago.ac.nz)

### RESUMEN

En los últimos años ha habido un gran crecimiento del potencial de las tecnologías "open source" (código abierto). En el 2005, la comunidad de código abierto geoespacial se consolida como una organización sin fines de lucro llamada la Open Geospatial Foundation (OSGeo). Colectivamente estos eventos han preparado el terreno para una época importante en el desarrollo de herramientas para aplicaciones geoespaciales. Este artículo describe el desarrollo de una herramienta que se originó a partir de estos eventos y fue desarrollada en la Universidad de Waterloo, Canadá, durante los últimos cuatro años.

Palabras clave: sistemas de apoyo a la toma de decisiones espaciales, sistemas de apoyo a las decisiones espacio-temporales, calidad de la educación, software de fuentes abiertas

### THE USE OF OPEN SOURCE SOFTWARE FOR SPATIAL DECISION SUPPORT

#### ABSTRACT

In recent years there has been a substantial growth in the potential of open source technologies. In 2005, the open source geospatial community formed a non-profit organisation called the Open Geospatial Foundation (OSGeo). Collectively, these events have prepared the way for an important period in the development of tools for geospatial applications. This article describes the development of a tool that arose from these events and was developed during the last four years at the University of Waterloo, Canada.

Keywords: spatial decision support systems (SDSS), spatio-temporal decision support systems (st-DSS), education quality, open source software

## 1. Introducción

En los últimos años ha habido un gran crecimiento del potencial de las tecnologías de "open source" (código abierto). El término open source (OS) fue introducido en 1998 durante una reunión de seis colaboradores, incluyendo a Eric Raymond, autor del texto fundamental de código abierto 'The Cathedral and the Bazaar' (Raymond, 2001) y en la que se debatió acerca de la decisión de Netscape de abrir al público el código fuente de su navegador de Internet. A pesar de que este evento es relativamente reciente, los orígenes del concepto de OS tienen una referencia mucho más antigua que se encuentra dentro de la historia del sistema operativo UNIX y otros software gratuitos para Internet en lo que muchos han descrito como la 'cultura hacker'.

La definición actual del término OS se ha *derivado* de las directrices de software libre de Debian (Perens, 1997), y las subsiguientes revisiones durante 1997 y 1998. La versión más actual (1.9) de esta definición (Open Source Initiative, 2006), demasiado extensa para reproducirla aquí, esencialmente gobierna los términos y condiciones bajo los cuales el software, incluyendo su código fuente, puede ser considerado 'libre' y puede ser distribuido "a" y usado por múltiples desarrolladores sin dejar de ser 'libre'. Esta definición y sus impactos sobre las comunidades de desarrolladores de software que la utilizan, así como en la industria de software propietario, han sido algunos de los desarrollos más importantes de esta última década dentro de la tecnología de la información.

Esta tendencia es más patente en Latinoamérica que en otros lugares. Aquí, varios gobiernos, incluyendo Chile, Brasil, Argentina, Perú y Venezuela han dado algunos pasos para crear una agenda Free/Libre/Open-Source Software (FLOSS) (Open Source Initiative, 2005). Entre éstos, el más notable es el caso de Perú, donde el gobierno decretó el uso de FLOSS por sus agencias en 2005 (por un voto de 61 a cero, con cinco abstenciones) (OSI, 2005; Asociación Peruana Software Libre, 2005).

Aunque la investigación de los orígenes y crecimiento del movimiento FLOSS valen la pena en sí mismos, este artículo se enfoca hacia el movimiento y debate de FLOSS y en el software libre geoespacial (OSG) (Schutzberg, 2006). La historia del software OSG es relativamente joven. Inicialmente se enfocaba a sólo un pequeño grupo de herramientas, la más notable el Geographic and Analysis Support System (GRASS) desarrollada por el United States Army Corps Of Engineers en 1982 y que se constituyó oficialmente como open source en 1999. En los Estados Unidos, el interés alrededor de GRASS evolucionó hasta llegar a la formación del Open Geospatial Consortium (OGC) en 1994. El OGC ha servido como el líder en la especificación de los estándares geoespaciales abiertos que muchas de las compañías de sistemas de información geográfica (SIG) más importantes, están empezando a reconocer y respetar en el desarrollo de su software.

En el 2005, la comunidad de OSG se consolidó como una organización sin ánimo de lucro llamada la Open Geospatial Foundation (OSGeo). OSGeo busca apoyar y construir software OSG de la mejor calidad posible, con la preferencia de fomentar una cultura de colaboración para el desarrollo de proyectos comunitarios. La formación de OSGeo fue impulsada por la decisión de AutoDesk Inc. de abrir la mayoría de los componentes de su software para crear mapas en la Web

(Map Guide). Sin embargo, es probable que la comunidad de desarrolladores de tecnologías OSG se unieran tarde o temprano, independientemente de la decisión de Auto Desk Inc.

Colectivamente estos eventos han preparado el terreno para una época importante en el desarrollo de herramientas FLOSS para aplicaciones geoespaciales (FOSS4G). Dichas herramientas sirven como fondo para el desarrollo de los programas académicos sobre software OSG y para la evolución de nuevas y útiles herramientas que tienden a ser orientadas a las *necesidades* y no a *respuestas del mercado* hacia oportunidades financieras.

Este artículo describe el desarrollo de una herramienta FOSS4G que surgió de tales necesidades y fue desarrollada en la Universidad de Waterloo, Canadá, durante los últimos cuatro años.

La herramienta llamada EduCal (la sigla significa 'educación de calidad' en castellano y 'education calculations' en inglés) fue desarrollada específicamente para facilitar la planificación descentralizada y la gestión de datos escolares a nivel de educación primaria<sup>1</sup> para el Ministerio de Educación Peruano (MINEDU) (en sus diferentes niveles e instancias descentralizadas, Organizaciones no-gubernamentales (ONG) e investigadores relacionados por su trabajo con el sector educativo en Perú, así como los otros usuarios que pueden tener acceso a la misma como los padres de familia). Tal estrategia fue diseñada para el uso de los datos rutinariamente recolectados de todas las escuelas estatales Peruanas por el MINEDU. La intención es agregarle 'valor' a estos datos al estructurarlos en una base de datos relacional y mostrarlos a través de Internet con herramientas OSG de código abierto de una manera flexible y con un modo de análisis multidimensional basado en escenarios.

El paquete de software EduCal le permite al MINEDU y a todos los actores involucrados en el sistema educativo peruano, gestionar un número grande de datos educativos espacialmente agregados. Específicamente, es posible seleccionar un pequeño grupo de escuelas individuales o diferentes áreas de interés usando la interfaz del mapa. Adicionalmente se pueden seleccionar indicadores de la base de datos de EduCal, incluyendo datos del Censo Nacional de Población y Vivienda, para llevar a cabo análisis espacial y temporal relacionado al rendimiento educativo.

Asimismo, se pueden identificar rápidamente escuelas y Unidades de Gestión Local (UGEL)<sup>2</sup> que se están quedando atrás en relación con los estándares nacionales y por tanto diagnosticar las causas de tales condiciones. Al mismo tiempo, la creación de escenarios dentro del software permite a los usuarios cambiar las escuelas/UGEL y/o los indicadores seleccionados para ver cómo estos pueden afectar los resultados.

Este artículo pretende ilustrar la segunda versión de EduCal. En las siguientes secciones se describirán las herramientas que se han desarrollado dentro del movimiento FOSS4G en relación a EduCal. Posteriormente se describirá ampliamente el diseño de EduCal y a continuación se realizará un resumen de los datos utilizados y los problemas relacionados con los mismos. Finalmente se concluirá con un ejemplo simple acerca del uso de la herramienta en el área metropolitana de Lima.

## 2. SIG de Código Abierto y la herramienta EduCal

Ramsay (2006) explicó los cuatro principios fundamentales que se deben cumplir para que un proyecto de fuentes abiertas (OS) tenga éxito. Específicamente, estos principios son:

1. El software que se desarrolla debe estar diseñado modularmente para que los miembros de la comunidad alrededor de la cual se está construyendo el software puedan contribuir a su desarrollo.
2. El software debe estar claramente y extensivamente documentado para que otros desarrolladores puedan programarlo de inmediato y agregar nuevas funciones.
3. El diseño central del software y el proceso de desarrollo deben ser transparentes para que el público tenga garantizado el acceso a todos sus aspectos a través de listas de distribución y foros on-line.
4. El equipo de trabajo central deber ser modular y transparente en si mismo con cierta fluidez en su composición para que los distintos miembros se puedan agregar y quitar de acuerdo a sus habilidades.

Si estos principios no están incorporados en la configuración de un proyecto de código abierto, la posibilidad de que éste tenga éxito disminuirá. En particular, disminuirá la posibilidad de lograr los objetivos y crear comunidades de programadores alrededor de proyectos. Por eso, la modularidad y buena documentación son dos ingredientes particularmente importantes. Sin embargo, estos ingredientes son difíciles de lograr, especialmente cuando la comunidad de miembros trabaja en distintos idiomas. Este hecho perjudica el desarrollo del tercer y, posiblemente, del cuarto principio. En estos casos, los idiomas que unen a la comunidad son los lenguajes de programación, dado que éstos son universales y no dependen ni de cultura ni de ubicación geográfica.

Además de satisfacer estos principios se requieren por lo menos tres componentes modulares para que un proyecto OSG que utiliza la Internet pueda brindar información a sus usuarios. Estos componentes incluyen:

1. Una base de datos relacional con capacidades espaciales.
2. Un servicio web de mapas (Web Map Service, WMS) o un servicio web de características (Web Feature Service, WFS) que conforme las normas de OGC y que permita exponer y obtener varios tipos de datos espaciales.
3. Un navegador que permita la interacción entre el usuario y la aplicación.

Típicamente, los dos primeros se encuentran en el mismo servidor mientras que la computadora del usuario se ubica en cualquier lugar de la Internet.

Para cumplir con los cuatro principios mencionados y satisfacer la arquitectura de

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, nº 8, p.23-42. ISSN: 1578-5157

---

componentes de un producto FOSS4G / OSG, los desarrolladores del proyecto EduCal inicialmente decidieron cual software OSG iban a utilizar para lograr la funcionalidad requerida y conseguir un producto final que facilitaría el análisis espacial y numérico de la calidad educativa utilizando Internet y la Web como medio de interacción con el usuario (Hall y Leahy, 2006).

Se seleccionó una combinación de herramientas OSG para arquitecturas de WMS para formar la base del proyecto EduCal. Específicamente, PostgreSQL fue seleccionada como base de datos relacional en conjunción con la herramienta de Postgre, PostGIS, para proveer las funciones espaciales. MapServer/Chameleon fueron seleccionados como el servidor de mapas e interfaz de usuario. La interfaz fue personalizada con programación propia usando MapScript PHP. Los componentes descritos aquí representan lo que es, posiblemente, el conjunto de productos OSG más común que se ha desarrollado. MapServer se puede considerar, sin duda alguna, el proyecto más exitoso de todos los proyectos OSG. Este éxito se debe en gran parte al desarrollo bajo los principios comentados anteriormente.

Sabiendo que existe una comunidad de usuarios con perfiles variados, el equipo de programación principal dedicó 100% de su tiempo a la mejora y mantenimiento del producto hasta obtener una herramienta bien documentada y estable. Cabe mencionar que MapServer es relativamente fácil de instalar y soporta la entrada y salida de una gran variedad de formatos de datos. Dichos factores convirtieron a MapServer en un verdadero competidor frente a herramientas de mapas comerciales que existen en el mercado (e.g. ESRI ArcIMG, GeoMedia Web Mapper y MapInfo MapExtreme).

La combinación de MapServer con PostGIS se ha convertido en una solución ideal para proyectos de apoyo en la toma de decisiones utilizando mapas de la Web, especialmente aquellos que requieren una arquitectura flexible y escalable.

Con la opción de controlar a MapServer y a PostGIS utilizando las interfaces de programación de aplicaciones de PHP, el proyecto EduCal pudo obtener considerable flexibilidad en los métodos de análisis que fueron implementados para el manejo de mapas así como en la gestión de la base de datos que supone la esencia fundamental del software.

Con la elección de los componentes ya decidida, cinco conceptos básicos tuvieron que ser cumplidos en la creación de la versión 2 de la herramienta, específicamente:

1. Conceptualmente tenía que existir una continuidad con la versión 1 para que los usuarios no se enfrentaran a un nuevo aprendizaje.
2. La versión 2 debía permitir que los usuarios accedieran a información de la base de datos de varios años. Sin agregar esta dimensión temporal para formar una especie de cubo tridimensional, no existiría una base sobre la cual determinar si, según el software, las escuelas han mejorado, empeorado o permanecido igual a través del tiempo.
3. La versión 2 debía permitir que los usuarios pudieran ejecutar operaciones de 'drill' (un término tomado del campo de la minería). Es decir, los usuarios tenían que poder hacer un 'drill down' (desagregar) a cada nivel espacial de la jerarquía del sistema educativo

Peruano (de áreas muy agregadas – Nacional - hasta las escuelas que son el nivel más desagregado) y 'drill across' entre distintos periodos de tiempo para el área seleccionada.

4. La versión 2 debía permitir que los usuarios pudieran subir en la jerarquía escolar (agregando – lo opuesto de 'drill down') a cada nivel espacial para examinar los cambios emergentes en la evaluación de la calidad de educación en relación a la escala espacial. Sin esto, no sería posible determinar hasta que punto los resultados en una determinada escala 'esconden' resultados con mayor variación en niveles más bajos.
5. Por último, la versión 2 debía permitir que los usuarios pudieran agregar sus propios indicadores a través de una interfaz intuitiva que permitiera añadir información nueva a la base de datos de EduCal. Esta función permitiría generar indicadores específicos de las condiciones locales sin depender de las encuestas nacionales del MINEDU.

Estos objetivos fueron obtenidos a través de su especificación en el diseño del software tal como se describe en la sección siguiente.

### 3. Una mirada amplia al diseño de EduCal

Cabe mencionar que sin un cuidadoso diseño no sería posible que una herramienta compleja como EduCal, fuera de uso productivo bajo las condiciones que se encuentran en un país como Perú. Como se ha indicado anteriormente, EduCal fue diseñado específicamente para ser utilizado por gestores de la educación en Perú pero se pretende que la arquitectura general pueda ser aplicada a la evaluación de la educación de cualquier país. Esta sección brinda una mirada amplia hacia dos aspectos del diseño de EduCal; la arquitectura del software y la lógica de flujo de datos dentro de la aplicación.

Como se ha indicado en la sección previa, EduCal se ha construido en base a la combinación de herramientas OS y OSG. Cada una de estas herramientas es fundamental para proveer la capacidad de crear mapas y almacenar y recuperar datos con una interfaz fácil de usar que mantiene vínculos entre cada función. Asimismo, cada herramienta fue integrada modularmente para efectuar una tarea específica y al combinarse en la manera correcta poder otorgar funciones espaciales potentes para ayudar en la toma de decisiones. La arquitectura de EduCal se puede ver en la [figura 1](#).

En el centro de la herramienta se sitúa PostgreSQL, un sistema de gestión de base de datos relacional (DBMS). PostgreSQL no tiene la capacidad de trabajar con datos espaciales, es decir, no puede ni almacenar ni procesar objetos espaciales (esencialmente puntos, líneas y polígonos). Para complementar la funcionalidad de PostgreSQL se ha añadido el módulo PostGIS que agrega soporte para objetos espaciales y un conjunto de funciones que permiten consultar la base de datos sobre estos mismos objetos.

Postgres/PostGIS constituyen el sistema de gestión de base de datos con capacidad espacial que se utiliza para soportar al resto del software. Por un lado proveen los datos al servidor de mapas

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, nº 8, p.23-42. ISSN: 1578-5157

---

(en este caso, MapServer). Por otro, lo hacen directamente a la aplicación. La aplicación utiliza los datos espaciales y los no espaciales mientras que el servidor de mapas utiliza la base de datos para generar mapas según los requerimientos de la aplicación.

La aplicación misma tiene dos propósitos. Por un lado contiene la lógica del flujo de la aplicación y también es responsable de generar la interfaz de usuario. Esta capa del sistema depende del ambiente Chameleon, el cual se ha construido usando el proyecto OSG MapServer como el servidor/generador de mapas. Siendo así, Chameleon funciona con todos los formatos de datos apoyados por MapServer a través de un archivo de tipo MAP. Estos componentes de EduCal forman el ambiente dentro del cual existe una serie de 'widgets' que permiten la ejecución de trozos de código PHP y JavaScript. Los dos módulos de PHP, `php_mapscript` y `php_postgres`, proveen de las bibliotecas necesarias para comunicarse con el servidor de mapas y con el DBMS respectivamente. La adición de una serie de 'widgets' programados especialmente para este proyecto por el equipo de EduCal en la Universidad de Waterloo, Canadá, ha extendido la funcionalidad de Chameleon substancialmente para proveer herramientas de exploración, selección y análisis más avanzados. Todas éstas están distribuidas de manera libre con el código fuente de EduCal.

Como en todas las aplicaciones Web, los usuarios interactúan con la aplicación a través de un navegador en su propio ordenador. Este ordenador se considera como 'cliente' y se comunica con el resto del sistema que se encuentra en el servidor. La combinación del servidor Web Apache y PHP es muy común. En la actualidad, más del 70% de los servidores Web utilizan Apache (SecuritySpace, 2006). A su vez `mod_php` es el módulo de Apache más popular (más de 40% de todos los servidores Web) (SecuritySpace, 2006).

Usando esta arquitectura, EduCal ofrece un proceso flexible e intuitivo para explorar y analizar los datos escolares. La lógica de flujo descrita en la [Figura 2](#) explica un proceso cíclico/iterativo para escoger los datos seleccionados y analizarlos. Ambos procesos son imprescindibles para que planificadores educativos y otros funcionarios que toman decisiones puedan entender y analizar problemas relacionados con la calidad de la educación sin necesitar tener conocimientos *a priori* (Leahy *et al.*, 2005). Estos problemas tienden en ocasiones a tener muchas dimensiones no bien definidas.

Queda implícito en el flujo descrito en la [Figura 2](#) que los analistas necesitan poder hacer selecciones de una manera rápida y sencilla. El analista sólo debe tener que pensar en 'qué' quiere seleccionar sin tener que pensar en 'cómo' hacerlo. Además, los resultados de cualquiera de los métodos de selección deben aparecer en menos de los 10 segundos identificados en la banda cognitiva de Newell (1990). Resultados que no satisfagan este requisito harían que el usuario perdiera su concentración y la confianza en la herramienta. Es decir, el proceso interactivo no debe ser interrumpido. Pruebas preliminares de EduCal versión 2 con los componentes y métodos descritos en las figuras 1, 2 y 3 han sido capaces de funcionar dentro del umbral propuesto por Newell. En la mayoría de los casos el software es capaz de responder en menos de cinco segundos al pedido de un usuario.

El usuario inicia un pedido al servidor cuando quiere explorar los datos. Esta exploración puede realizarse de dos formas; el usuario analiza los datos directamente o define su selección con

una de las tres acciones mostradas a la izquierda de la [Figura 2](#). De cualquier manera, los resultados de un análisis no son el producto final de este proceso. Se supone que los mapas y las tablas sirven para guiar al usuario en su elección de 'cuál' acción es la más apropiada según los resultados preliminares. Al ver estos resultados, el usuario vuelve a su selección actual y toma una de las acciones según la situación. Cada acción brinda una nueva actualización de la selección y constituye el principio de una nueva ronda de análisis. Durante cada ciclo existe la opción de guardar una copia de resultados interesantes para poder compartirlos con otros usuarios, analizar de otra manera o simplemente para volver al análisis en otro momento.

Conceptualmente, cada una de las acciones cambia la selección a lo largo de una dimensión distinta. Estas dimensiones y sus acciones correspondientes se explican, junto al modelo para el manejo de los datos, en la sección siguiente.

#### 4. Datos educativos y los problemas con los datos en el Perú

Los datos educativos se pueden dividir a lo largo de las dimensiones de espacio, tiempo y atributos (indicadores). Además, cada una de estas dimensiones tiene su propia multidimensionalidad. Por ejemplo, la dimensión espacial se puede agregar y desagregar teniendo en cuenta los niveles jerárquicos correspondientes del sistema educativo. En el nivel más bajo, los datos están vinculados con los objetos espaciales más pequeños (escuelas y 'centros poblados' y a las múltiples instancias de tiempo -años- para formar regiones espacio-temporales) (Leahy *et al.*, 2005). Esta organización de datos en forma de cubo se muestra en la [Figura 3](#).

Usando este modelo, cada acción se puede conceptualizar como un corte o tajada del cubo a lo largo de una de las dimensiones. Por ejemplo, la selección de un área usando su nombre tomaría una tajada del cubo en la [Figura 3](#) a lo largo del 'eje-y', una selección usando atributos tomaría una tajada a lo largo del 'eje-x', y las selecciones que involucran el tiempo tomarían una tajada a lo largo del 'eje-z'. Las acciones 'drill down' y 'roll up' son diferentes porque no cambian las dimensiones seleccionadas. En cambio, estas acciones agregan y desagregan las dimensiones ya seleccionadas. Es importante indicar que estas acciones se pueden ejecutar independientemente de las otras y entonces es posible que una selección se defina a lo largo de cualquier número de ejes al mismo tiempo.

Los datos que forman este cubo de EduCal se han obtenido principalmente de dos fuentes: el Ministerio de Educación Peruano (MINEDU) y el Instituto Nacional de Estadística e Informática (INEI). Los datos educativos provienen del Censo Escolar 2002, 2004 y del Censo de talla 1999, ambos proporcionados y recolectados por el MINEDU. Los datos contextuales provienen del Censo Nacional de Población y Vivienda 1993 que fue adquirido del INEI. Asimismo, se han recogido algunos datos a través de trabajo de campo realizado por el equipo de la ONG Alternativa.

Para el caso de los datos educativos recopilados por el MINEDU, el acceso a los mismos ha sido relativamente sencillo pero su incorporación a una base de datos relacional ha sido complicada debido a la estructura de los datos adquiridos. En algunos casos, la definición de los campos no guarda relación entre cada año (2002 - 2004 - 2005), debiéndose emplear muchas horas de trabajo para asegurar la continuidad y confianza de la información recopilada. La coherencia de los datos es

imprescindible para facilitar la comparación entre un censo escolar y otro.

Adicionalmente a las complicaciones señaladas, los funcionarios no necesariamente pueden garantizar la confianza absoluta de la información recopilada. Por ejemplo, existen casos de escuelas que inflan el número de alumnos para garantizar su funcionamiento y garantizar el contrato de personal o la permanencia del mismo, en caso de ser estable, para que no se le asigne a otra zona que sí puede necesitar docentes. Asimismo, muchas escuelas privadas no entregan información o sólo lo hacen parcialmente.

La información educativa que es proporcionada por estas fuentes de información es de carácter fundamentalmente cuantitativa con datos de estudiantes (sexo, número de promocionados, de repetidores, de matriculados, etc.), docentes (sexo, condición laboral, certificación, entre otras), infraestructura educativa (número de aulas, secciones, mobiliario, laboratorios, bibliotecas, computadoras, etc.), etc. Al nivel de las escuelas, no se cuenta con fuentes de información sobre los procesos de enseñanza / aprendizaje que se dan en el aula, evaluaciones de comprensión de lectura, etc., o información contextual del niño que permita conocer datos sobre los padres, el grado de instrucción de los mismos, etc. Toda esta información se recoge en la Ficha Única del Estudiante pero no se encuentra digitalizada.

Como parte de la información contextual educativa también se ha incluido el Censo de Talla del año 1999. Si bien este censo tiene siete años de antigüedad y no guarda relación con la población escolar del censo 2002 o 2004, permite tener una idea del nivel nutricional promedio de la mayoría de la población escolar. Junto al censo escolar, el censo de talla completa la parte principal de los datos educativos de EduCal.

Para complementar los datos educativos, se han agregado datos de contexto socio-económico. Esta información contextual tiene una antigüedad de más de 10 años pues tiene como fuente el Censo de Población y Vivienda de 1993. Estos datos existen a todos los niveles de agregación desde el nivel mínimo de manzana proporcionado por el INEI. Ha sido necesario construir indicadores según especificaciones propias ya que las variables proporcionadas por el INEI no eran relevantes al proceso educativo. Se incurrió un costo en el procesado de estas variables ya que los datos al nivel de la ficha censal no están disponibles al público por razones de confidencialidad. En el momento de recopilar los datos no existían datos socio-económicos más recientes.

Sin embargo, en el año 2006 se desarrolló un nuevo Censo de Población y Vivienda. Este censo aún no se ha incluido dentro de la herramienta EduCal, pero existe interés en hacerlo. Se anticipan algunas complicaciones pues el menor nivel de desagregación en el que se recopila información de este censo es el distrito. También cabe señalar que, al igual que en el caso de los datos educativos, el número de variables recopiladas es sustancialmente menor que aquel del Censo de 1993. Por este motivo, algunos de los indicadores no se podrán comparar a través del tiempo.

El INEI también ha sido el proveedor de datos espaciales con la cartografía actualizada en el 2001. El MINEDU ha aportado gratuitamente cartografía actualizada del INEI a nivel de Centro Poblado y Escuelas. Ambas fuentes proveen al público la cartografía en formato de Shapefile. Este formato propietario se ha podido incorporar en la base de datos de EduCal sin mayores

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

inconvenientes.

Sin embargo, uno de los problemas detectados en los datos espaciales es la ausencia de metadatos. La falta de esta información complementaria no permite saber cuando fue creada la cartografía o en qué tipo de proyección geográfica se encuentran los datos. Ambas carencias dificultan la combinación de las dos fuentes de información usando el mismo juego de coordenadas espaciales.

Como se ha señalado, ninguna de las fuentes de datos ha podido ser incorporada sin dificultades. Se concluye que existe una falta de experiencia por las agencias peruanas en la gestión de sus datos, lo cual ha hecho laboriosa la creación de la base de datos que se encuentra en EduCal. No obstante, a pesar de todos los problemas con la adquisición y manipulación de los datos, el resultado final ha sido un almacenamiento de los mismos bastante comprensible y consistente.

## 5. Ejercicio aplicado utilizando EduCal

La existencia de datos acerca de varios aspectos educativos y la inclusión de estos mismos agregados para cada nivel de la jerarquía del sistema escolar, hacen que EduCal sea útil para distintos tipos de usuarios. EduCal facilita la exploración y análisis de los datos de los diferentes censos y encuestas escolares de una manera pertinente para cada actor del sistema educativo, incluyendo a los estudiantes, padres, maestros y funcionarios del MINEDU locales, regionales y nacionales.

En el ejemplo siguiente, se ha asumido el rol de un funcionario del nivel regional del MINEDU en Lima Metropolitana. Para un gestor al nivel regional, es importante tener un buen conocimiento de lo que está pasando en los distintos barrios y escuelas dentro su área de responsabilidad. Además, es importante tener la habilidad de dirigir recursos hacia las escuelas que están teniendo las mayores dificultades para poder maximizar el impacto positivo a la calidad de educación recibida por los estudiantes.

En este ejemplo, la Unidad de Gestión Educativa Local 02 (UGEL 02) ha sido designada como área de interés. La UGEL 02 tiene cuatro distritos y 138 escuelas. Aunque EduCal permite computar índices de calidad de educación basados en indicadores múltiples, para este ejemplo se utilizará la *Tasa de Promoción de Estudiantes* (TPE). TPE se considera como un indicador importante para cuantificar el éxito de las escuelas en ayudar a los estudiantes para avanzar dentro del sistema educativo en el número de años adecuados.

La UGEL 02 se puede seleccionar dentro de EduCal haciendo una consulta con su número o seleccionándola en el mapa. Como resultado tenemos que el TPE para la UGEL 02 en 2004 es 0.86 (86% de los estudiantes fueron aprobados para pasar al siguiente grado) mientras que en 2002 este número era esencialmente igual, 0.87. Para obtener una perspectiva de la calidad de la UGEL 02, se pueden seleccionar todas las UGEL de Lima y calcular algunas estadísticas a este nivel de agregación, como muestra la [Tabla 1](#).

Según estos resultados, la UGEL 02 está al mismo nivel que las otras UGEL en el área. Sin

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

embargo, al extender la selección a todas las UGEL de Perú se puede ver que el TPE nacional promedio en 2004 fue 0.74. Es decir, todas las UGEL de Lima están aproximadamente un 10% más altas que el promedio en el resto del país.

Sin embargo, al ver el TPE agregado al nivel de la UGEL parece que el rendimiento de los estudiantes es bueno. Al inspeccionar el TPE a nivel de los distritos se puede ver que existen variaciones *dentro* de la UGEL. En la [Figura 4](#), se aprecia que en el distrito de Los Olivos hay una tasa de promoción de  $\leq 0.85$  mientras que el distrito del Rimac está por delante del resto. Parece ser que a nivel de distritos, hasta el que podemos considerar el peor distrito, tiene una tasa de promoción aceptable. El desagregar los datos a nivel de escuelas permite ver resultados distintos.

La [Figura 5](#) muestra que existe cierta variación entre las escuelas de la UGEL 02. Los datos agregados presentados antes ocultaban la existencia de escuelas con problemas dentro del área de estudio. A través de operaciones 'drill up' y 'roll down' del cubo de datos de EduCal, es posible llegar a un entendimiento local y global de la situación *dentro* de la UGEL. Esta capacidad de EduCal de poder cambiar rápidamente y fácilmente entre los distintos niveles de agregación espacial es una parte indispensable del análisis de calidad de educación.

El establecer que existen escuelas con una tasa de promoción baja nos lleva a identificar exactamente cuáles son las escuelas problemáticas y si es posible, establecer las causas de los problemas. A través de la formulación de consultas, las escuelas con una TPE de menos de 0.75 (3/4 de los estudiantes) pueden ser seleccionadas ([Figura 6](#)).

De las 138 escuelas quedan seleccionadas 25. Al ver estas escuelas en forma de tabla y ordenas por los valores del TPE ([Figura 6](#)) se puede identificar que *2015 Manuel González Prada (ID 4362610)* es la escuela con la tasa más baja. Esta escuela tiene una tasa de promoción de menos de 20% en 2004. En 2002, la TPE también era baja, aunque bastante mejor. Se puede deducir que en los dos años algo ha cambiado y ocasionado que la escuela empeore su rendimiento.

Después de identificar las peores 25 escuelas en la UGEL y a su vez la peor entre éstas, el próximo ciclo de análisis puede tomar varias direcciones. Podemos entonces extender o redefinir la selección con la ayuda del formulador de consultas para responder preguntas y avanzar con el proceso de descubrimiento.

Por ejemplo, es posible establecer si la tasa de promoción de *Manuel González Prada* es la norma en el área inmediata o si el rendimiento en la escuela es una anomalía. Para poder hacer esto, se puede usar de nuevo el formulador de consultas, esta vez para seleccionar todas las escuelas dentro de 2 km de *Manuel González Prada*. La elección de 2 km para definir el área inmediata de la escuela es arbitraria pero suficiente para dar una idea de las condiciones alrededor de la escuela. Por supuesto, cualquier otra métrica de proximidad podría ser seleccionada igual de fácilmente. Existen 30 escuelas dentro de un radio de 2 km y tienen una TPE con un máximo de 1, una media de 0.829 y mínimo de 0.19 ([Figura 7](#)). Estos números demuestran claramente que la causa de la TPE extremadamente baja de *Manuel González Prada* no está afectando a las escuelas en su entorno, por lo tanto no es un problema de la comunidad entera, ya que aparenta que la causa de los problemas en *Manuel González Prada* no se encuentra en su área inmediata. Vale la pena, por tanto, explorar qué escuelas tienen características parecidas. Es decir, ¿Qué otras escuelas en la UGEL 02 tienen una

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

TPE de menos de 0.75 y han empeorado por lo menos 0.2 desde 2002? Esta consulta se puede hacer fácilmente con el formulador de consultas de EduCal, como se demuestra en la [Figura 8](#). Aunque todas las escuelas con una TPE de menos de 0.75 llaman la atención, las que han empeorado desde el año anterior son particularmente alarmantes. El resultado de esta consulta son las cinco escuelas en la [Figura 9](#).

Vale la pena investigar más estas cinco escuelas, de las cuales tres están en el distrito Los Olivos y dos en el distrito de San Martín de Porres. Para empezar, otros indicadores se podrían seleccionar para buscar factores que tengan una correlación. También, ya que el grupo de escuelas identificadas es pequeño, se podría contactar a los directores de cada escuela directamente. Los directores seguramente tienen algunas ideas de cuáles pueden llegar a ser las causas de sus problemas. Alternativamente se podría usar el formulador de consultas para identificar las escuelas que han mejorado (por ejemplo, una mejora superior a 0.2) para identificar las condiciones que brindan una mejora en la TPE.

En este ejemplo, EduCal ha ayudado al planificador escolar de la Región de Lima Metropolitana a demostrar rápidamente que, a pesar de que la TPE al nivel nacional es aceptable en la UGEL, existen escuelas que sí están teniendo problemas. A través de una serie de operaciones 'drilling' y consultas, un conjunto pequeño y manejable de cinco escuelas fue identificado como parte de una investigación más profunda. Además, se estableció que el problema no es ni sistémico ni concentrado en sólo un área, permitiendo que se dirijan recursos a las escuelas individuales y no al sistema completo. Este ejemplo sólo ha inspeccionado un aspecto pequeño de lo que seguramente es un tema mucho más complejo. Sin embargo, las herramientas y las técnicas demostradas aquí se pueden aplicar para explorar muchos de los otros aspectos del sistema educativo local.

## 6. Conclusión

En base a la discusión presentada en este artículo, es evidente que el software FOSS4G representa una solución viable y práctica para muchos de los usos de los datos espaciales. En este caso dos proyectos OSG, es decir MapServer Chameleon y PostGIS, y un proyecto de código libre, PostgreSQL, fueron utilizados con código personalizado en PHP (con mapscript) y Javascript para realizar un sistema de alto rendimiento en el análisis de la calidad de la educación en Perú.

Los resultados de este proyecto han sido muy satisfactorios, en el sentido de que la herramienta es escalable para bases de datos grandes (como las 33,000 escuelas estatales de Perú) y que se puede personalizar de una manera relativamente simple para otro país con un sistema educativo jerárquico. Sin embargo, los esfuerzos que han sido necesarios para llegar al producto final no se deben minimizar. El software fue diseñado e implementado por cuatro desarrolladores en Canadá, por lo tanto, la 'infraestructura' del proyecto varía con respecto a los orígenes de los proyectos de código abierto convencionales tratados en la sección 2. Aunque hubo una cooperación importante e intensa entre los programadores, facilitadores (la ONG Alternativa en Lima) y los usuarios del sistema educativo peruano, el proyecto EduCal sigue siendo un proyecto OSG sin mucho apoyo. Una vez que los estudiantes de maestría que han desarrollado la herramienta en Canadá avancen en sus carreras fuera de la Universidad, el proyecto perderá impulso, al menos que nuevos programadores se incorporen al proyecto, sea en Canadá, Perú o en otra parte.

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

Ya que la herramienta es 'abierta' en el sentido de que todo el código fuente está disponible para que cualquiera pueda obtenerlo y modificarlo, esperamos que otros en Latinoamérica la tomen, la modifiquen y la usen dentro de su contexto. También nos gustaría comentar que el funcionamiento demostrado en la sección 5 es sólo una pequeña parte de lo que se puede alcanzar con el software en su estado actual. Con tiempo, recursos y convicción, cualquier cosa es posible dentro del marco de OS(G) dentro del cual se creó EduCal. Es decir, el funcionamiento se puede extender para incluir el uso de estadísticas espaciales y optimizar la gestión de mapas y de otros aspectos inherentes a los datos espaciales.

La presencia de la herramienta EduCal realmente converge con las leyes últimamente dictadas que establecen la descentralización de la responsabilidad de planificación, transparencia en el uso y acceso a datos coleccionados por el Estado, así como el uso de productos OS(G) por el gobierno. La idea básica de esta herramienta es devolver los datos a las escuelas a través de Internet, dentro de un paquete de software fácil de usar que ayude en la planificación y que tenga un robusto y adecuado funcionamiento en un país como Perú.

Este mismo método se puede extender a muchos otros campos. De este modo, la necesidad previa de adquirir software caro para gestionar bases de datos y datos espaciales puede dejar de ser un problema porque los usuarios pueden buscar más allá de las herramientas propietarias comerciales que se encuentran en el mercado SIG actual.

A medida que los proyectos OSG sigan evolucionando y volviéndose más sofisticados en su funcionamiento se convertirán también en más fáciles de instalar y usar. Hay una presencia fuerte de desarrolladores de OSG en Latinoamérica, especialmente en Brasil, y es probable que esta presencia siga creciendo y prosperando cuando se expanda el conocimiento de lo que es posible con creatividad y trabajo.

### Referencias bibliográficas

- Asociación Peruana Software Libre (2005): *Congreso Peruano aprueba ley a favor de software libre*, <http://apesol.org.pe/news/197>, Last Accessed 20th Dec, 2006.
- Hall, G. B. y Leahy, M.G. (2006): "Internet-based Spatial Decision Support using Open Source Tools", *Collaborative Geographic Information Systems*, edited by Shivanand Balram and Suzana Dragičević, Hershey, PA: Idea Group.
- Leahy, M.G., Hall, G.B., y Alperin, J.P. (2005): "Enabling space-time data mining of education databases using Open Source GIS and Internet mapping Tools", *Proceedings of Open Source Geospatial 2005*. Minneapolis, Minnesota, 18<sup>th</sup> June.
- Newell, A. (1990): *Unified theories of cognition*, Harvard University Press, Cambridge.
- Open Source Initiative (2005): *Peruvian Congressman refutes Microsoft's "Fear, Uncertainty and Doubt"*. (F.U.D.) Concerning free and open source software, [http://www.opensource.org/docs/peru\\_and\\_ms.php](http://www.opensource.org/docs/peru_and_ms.php), Last Accessed Dec 20<sup>th</sup>, 2006.
- Open Source Initiative, (2006): *The Open Source definition*, <http://www.opensource.org/docs/definition.php>, Last Accessed Dec 19<sup>th</sup>, 2006.
- Perens, B. (1997): *Debian Free Software Guidelines*, [http://www.debian.org/social\\_contract#guidelines](http://www.debian.org/social_contract#guidelines), Last Accessed Dec 19<sup>th</sup>, 2006.
- Ramsay, Paul, (2006): *The state of Open Source GIS*, Refrations Research Inc. White Paper,

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): “El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones”, *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

[http://www.refractions.net/white\\_papers/oss\\_briefing/2006-06-OSS-Briefing.pdf](http://www.refractions.net/white_papers/oss_briefing/2006-06-OSS-Briefing.pdf), Last Accessed Dec 19<sup>th</sup>, 2006.

Raymond, E. (2001): *The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary*, Revised Edition, Sebastopol, Ca., O’Reilly and Associates.

Schutzberg, Adena (2006): “Open Source and Open Standards in Geospatial Technologies: two kinds of open come together”, *Geoinformatics*,

<http://www.geoinformatics.com/asp/default.asp?t=article&newsid=2265>, Last Accessed Dec 20<sup>th</sup>, 2006

SecuritySpace (2006): [http://www.securityspace.com/s\\_survey/data/200611/index.html](http://www.securityspace.com/s_survey/data/200611/index.html) and [http://www.securityspace.com/s\\_survey/data/man.200611/apachemods.html](http://www.securityspace.com/s_survey/data/man.200611/apachemods.html), Pages Last Accessed Dec 19<sup>th</sup>, 2006.

## TABLAS

**Tabla 1. Tasa de promoción de estudiantes de todas las UGELEs de Lima.**

	<b>2002</b>	<b>2004</b>
<b>Maximum</b>	0.92	0.87
<b>Median</b>	0.88	0.85
<b>Minimum</b>	0.85	0.84

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

## FIGURAS

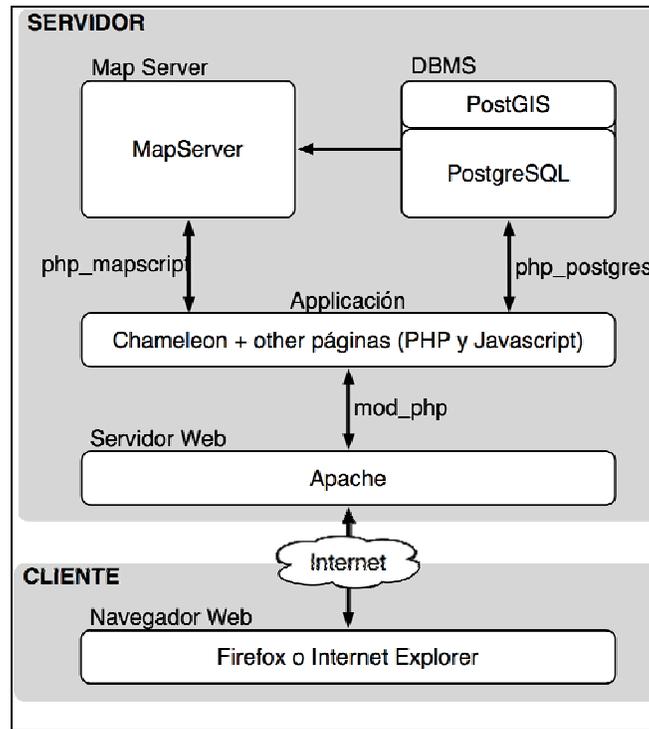


Figura 1. Arquitectura de EduCal.

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

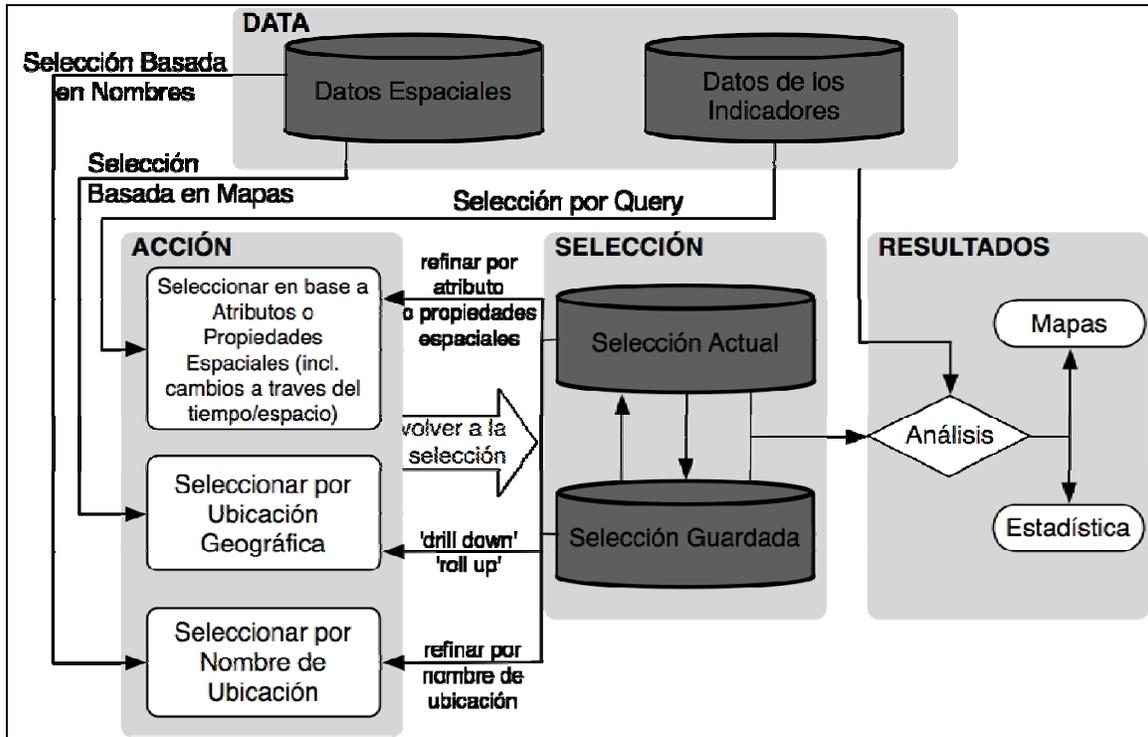


Figura 2. Lógica de flujo de la aplicación.

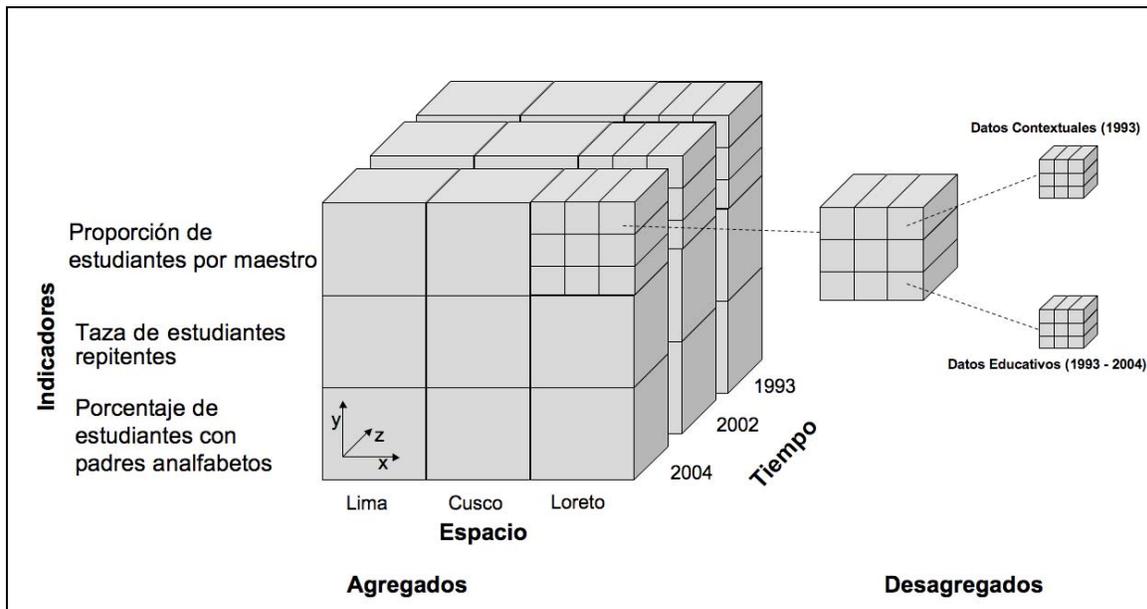


Figura 3. Modelo de cubo.

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, nº 8, p.23-42. ISSN: 1578-5157

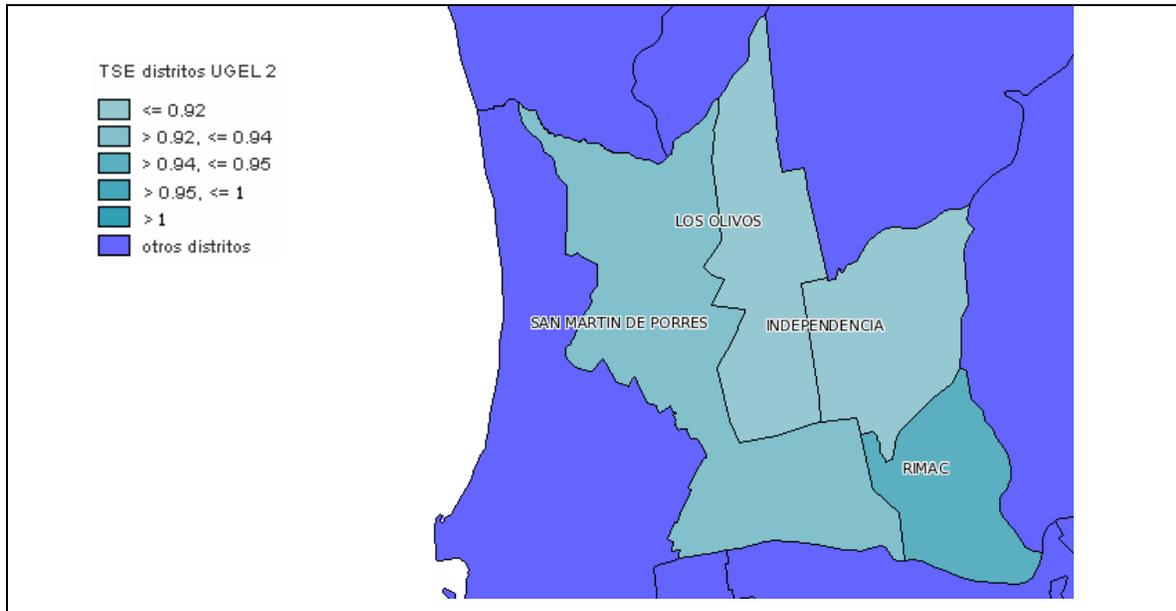


Figura 4. Tasa de promoción de estudiante de los distritos de la UGEL 02.

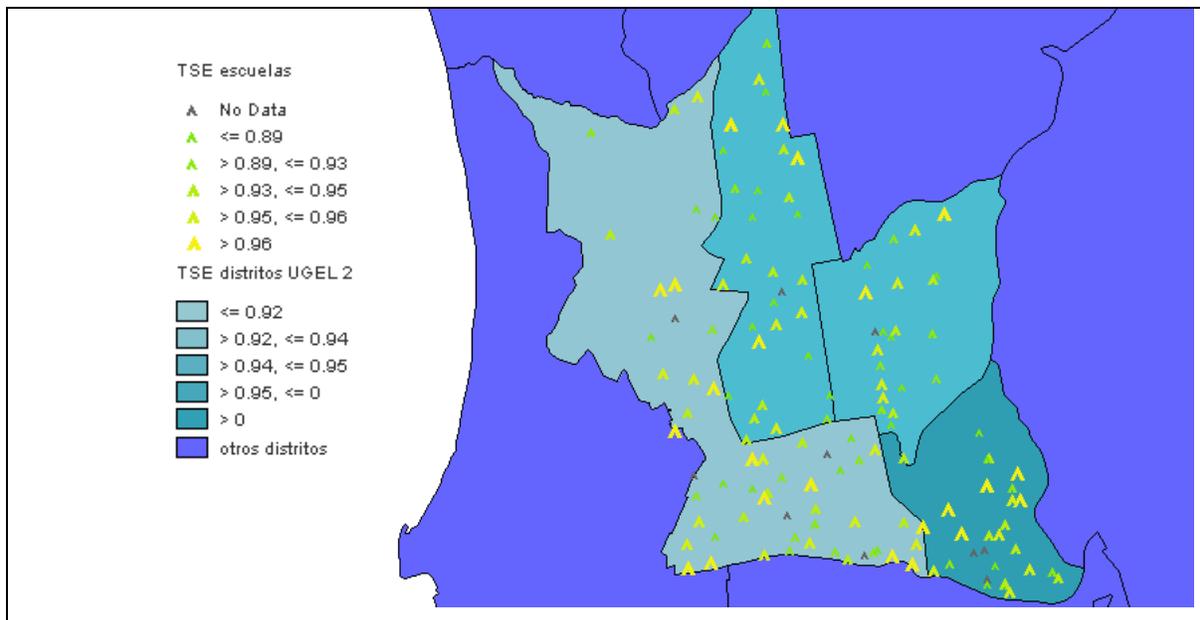


Figura 5. Tasa de promoción de estudiantes de las escuelas.

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, nº 8, p.23-42. ISSN: 1578-5157

Layer Name: Map Selection      Sort: Overall student promotion ra

School Identifier	School Name	Overall student promotion rate (year=2002)	Overall student promotion rate (year=2004)
4362610	2015 Manuel Gonzales Prada	0.57	0.19
8552470	Nuevo Peru	0.87	0.47
8210250	2034 Virgen De Fatima	0.87	0.52
4341340	3043 Ramon Castilla	0.89	0.53
6636820	3093	null	0.56
4366340	2058	null	0.6

Total Pages: 1  
Current Page: 1      Previous    Next    Download    Close

Figura 6. Resultados en forma de tabla y ordenados por TPE.

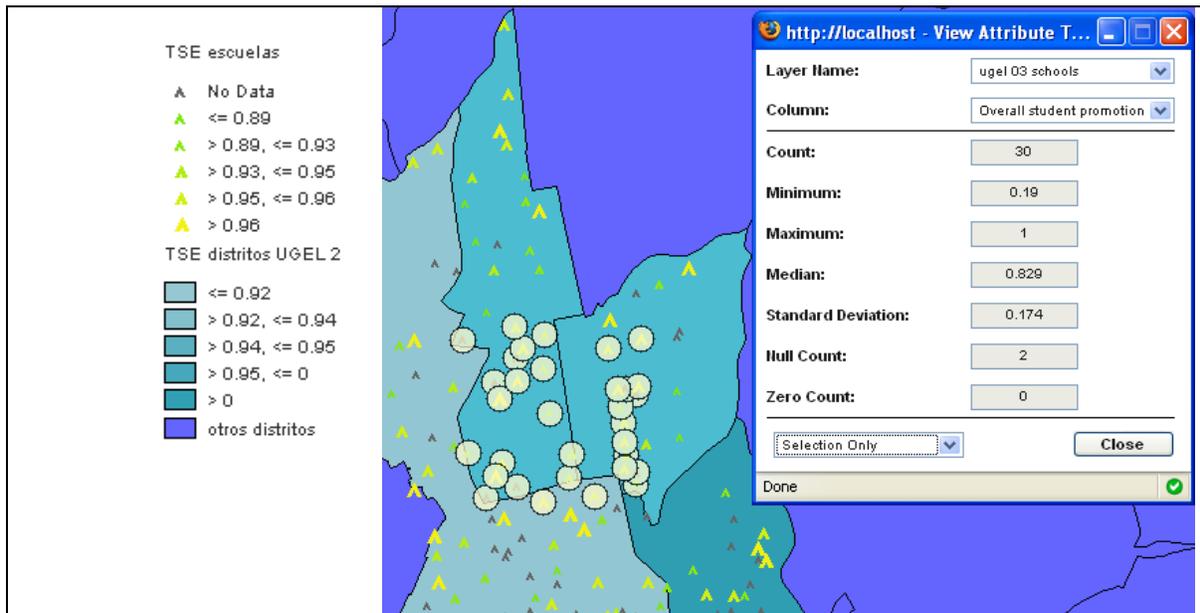


Figura 7. Escuelas dentro de un radio de 2 km de Manuel González Prada.

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

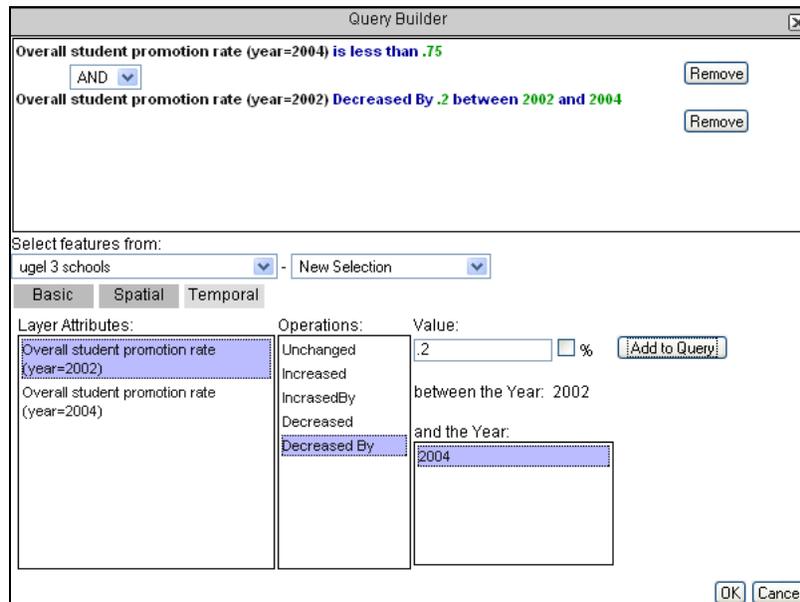


Figura 8. Formulador de consultas.

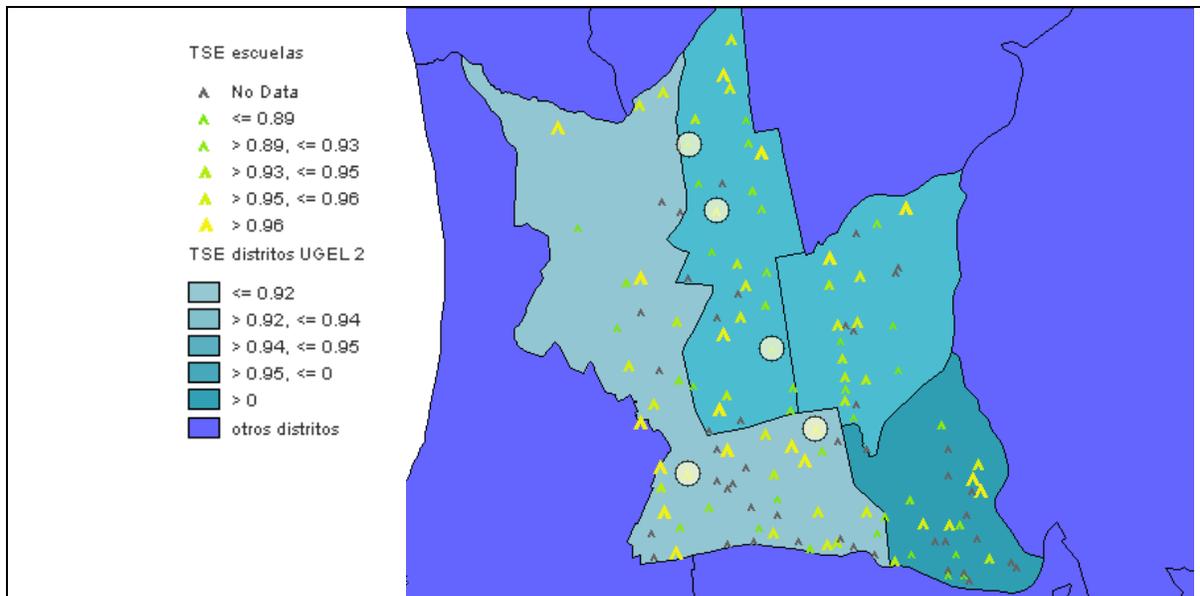


Figura 9. Resultados de la consulta.

<sup>1</sup> En el Sistema Educativo Peruano existen varios niveles educativos: Educación Inicial (comprende la

Hall, G. B., Alperin, J. P. y Kerrigan León, S. (2008): "El uso de Internet con software libre y fuentes espaciales abiertas para colaborar en la toma de decisiones", *GeoFocus (Informes y comentarios)*, n° 8, p.23-42. ISSN: 1578-5157

---

primera infancia (3 a 5 años), Educación Primaria compuesto por 6 grados (años), Educación Secundaria compuesto por 5 grados. Educación Superior que puede ser técnica o universitaria. Por ley, el Estado garantiza que todo niño / niña peruano debe cumplir el 2° y 3er nivel como mínimo.

<sup>2</sup> La jerarquía administrativa del MINEDU establece varios niveles de gestión administrativa y pedagógica. En el nivel base se encuentra la Escuela o Institución Educativa (IE) y en el nivel inmediatamente superior se sitúa la Unidad de Gestión Local (UGEL) que coincide con el ámbito de las Provincias, salvo para Lima Metropolitana que cuenta con 7 UGELes. A continuación viene el nivel de Dirección Regional de Educación (DRE) que coincide con las Regiones o Departamentos. Finalmente la instancia nacional o el MINEDU. Desde la Instancia Nacional hasta la Intermedia previa y el IE, no sólo tienen competencias administrativas y de gestión sino también fundamentalmente pedagógicas, siendo por tanto organismos de orientación.