
DESARROLLO DE UNA SOLUCIÓN GEOBI PARA LA OBSERVACIÓN Y ANÁLISIS DE LA INFORMACIÓN DE CENSO EN CHILE

SEBASTIÁN ARACENA AGUIRRE, RICARDO VALDIVIA PINTO, MÓNICA NAVARRETE ÁLVAREZ y ROBERT CORNEJO YÁÑEZ

RESUMEN

Se describe una solución de Inteligencia de Negocios para los censos realizados en Chile en los años 1992 y 2002, con el objetivo de integrar la información de ambos periodos para facilitar la observación y análisis espacio-temporal de los datos. La solución involucró la implementación de un Data Mart que incluye información geoespacial (GeoBI), lo que permite el uso de tecnología SIG y herramientas espaciales de procesamiento analítico en línea (SOLAP) para la visualización de la información. Para ampliar el ámbito de difusión de esta información, se ha implementado un servidor de mapas web 'Geo Análisis Censo',

a través del cual es posible interactuar dinámicamente con la información de población, considerando el uso de cubos, cartografías y dashboards. La alternativa existente en el mercado es REDATAM, herramienta ampliamente utilizada en Latinoamérica para gestionar la información de censos. Sin embargo, ésta no permite realizar consultas integradas entre censos de diferentes años, ni interactuar con la visualización cartográfica de los datos. Por tanto, este trabajo contribuye a la observación y análisis de la geoinformación censal con herramientas de libre acceso que posibilitan un análisis integrado más flexible y potente.

Los censos de población, vivienda y hogares constituyen una de las fuentes de información estadística más importante a nivel país, cuyos resultados suministran los antecedentes necesarios para el diseño e implementación de políticas y programas públicos, y para la decisión de inversión y estudios privados (INE, 2002). Realizado por lo común cada 10 años por los Institutos Nacionales de Estadística en cada país, es posible descargar la información desde sus páginas web institucionales. El Centro Latinoamericano y Caribeño de Demografía (CELADE), División de Población de la Comisión

Económica para América Latina y el Caribe (CEPAL), ofrece un gestor de datos censales conocido como REDATAM y que posee funcionalidades básicas de un sistema de información geográfica (SIG) y de acceso a los microdatos censales de población y vivienda, control migratorio y agropecuario, entre otros para varios países (CELADE, 2013).

En Chile se utiliza REDATAM+SP en su versión *desktop*, donde se pueden realizar consultas de desagregación espacial por región a nivel de manzana, mientras que con la versión *Web Server* se puede consultar hasta nivel de comuna (CELADE, 2004). Sin embargo,

pese a su gran utilidad, esta solución no permite realizar comparaciones temporales entre censos de diferentes años en una misma consulta. Por otro lado, la información cartográfica en la versión *Web Server* sólo puede ser obtenida en archivos de imagen (de formato PNG) impidiendo la interactividad con los datos que se muestran.

Aplicaciones tecnológicas que ofrecen el desarrollo de plataformas de mapas Webs conocidos (SIIGSA, 2014) como sistema de información territorial (SIT) y sistemas de consultas estadístico territorial (ICET), otorgan la posibilidad de generar cartografía en distintos niveles de

PALABRAS CLAVE / Data Mart / ETL / GeoBI / Inteligencia de Negocios / SIG / SOLAP /

Recibido: 21/01/2014. Modificado: 05/08/2014. Aceptado: 13/08/2014.

Sebastián Aracena Aguirre. Ingeniero Civil en Computación e Informática y, Magister en Ingeniería de Software, Universidad de Tarapacá (UTA), Chile. Dirección: Área de Computación e Informática, EUIIIS, UTA. 18 de Septiembre 2222. Arica, Chile. e-mail: sebastian.aracena@alumnos.uta.cl

Ricardo Valdivia Pinto. Ingeniero Civil en Informática, Universidad de Santiago de Chile. Doctor en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile. Profesor, EUIIIS UTA, Chile. e-mail: rvaldivi@uta.cl

Mónica Navarrete Álvarez. Ingeniero Comercial, UTA, Chile. Doctora en Modelización Económica Aplicada, Universidad Autónoma de Madrid, España. Profesora, UTA, Chile. e-mail: mnavarre@uta.cl

Robert Cornejo Yáñez. Ingeniero Civil en Computación e Informática y Magister en Ingeniería de Software, UTA, Chile. Investigador y Analista, UTA, Chile. e-mail: robert.cornejo.uta@gmail.com

desagregación territorial (región y comuna), añadiendo funcionalidades dinámicas para la interacción de la información a través de la cartografía y para generar reportes (SIT, 2014). Un ejemplo de esto es la plataforma utilizada por la Oficina de Estudio y Políticas Agrarias (ODEPA) la cual entrega consultas de encuestas de estadísticas continuas y del Censo Agropecuario con una amplia información territorial y que permite la interacción de los datos (ODEPA, 2014). Sin embargo, estas soluciones no tienen arquitectura de SOLAP, (por las siglas de *spatial on-line analytical processing*) y por eso no permiten cruzamiento de datos de distintas variables; por tanto, para cada variable existe una cartografía que asocia un indicador de tipo social (SIT) o variables del Censo Agropecuario (ICET).

En este trabajo se muestra el desarrollo de una solución *Web Server* que integra tecnología SIG e Inteligencia de Negocios para tratar datos mediante el acceso a un repositorio integrado de información de datos pertenecientes a censos de distintos periodos y con posibilidades de obtener desagregaciones microterritoriales menores a la comuna, como es el caso de los distritos. Estos datos pueden ser georeferenciados y analizados, posteriormente, con herramientas de inteligencia de negocios geoespacial (GeoBI, de *geospatial bussines intelligence*; Wickramasuriya *et al.*, 2013).

Aquí se describe el proceso requerido para gestionar los datos de censo desde archivos de formato SPSS (IBM SPSS Statistics) a un modelo de datos que describe la información del censo. El proceso de extracción, transformación y carga (ETL) a partir de los archivos SPSS, se construyó a partir del modelo requerido. Se utilizaron bases de datos intermedias (conocidas como *staging area*) para el proceso de limpieza de los datos. El trabajo se organiza en un marco teórico, la metodología de desarrollo de la solución (que involucra la construcción de un *Data Mart* y la implementación de un servidor de mapas con soporte SOLAP), un análisis de los resultados obtenidos, discusión y conclusiones.

Marco Teórico

Para interrelacionar los datos censales pertenecientes a diferentes fuentes de información, se recurre a un repositorio denominado *Data Warehouse* o fuente secundaria de información (pues no generan datos por sí mismos, sino que son alimentados desde sistemas existentes) el cual es una colección de datos obtenida a partir de datos transaccionales, específicamente estructurada para realizar

análisis de información (Kimball, 1992) y que involucra un alto volumen de información. Para que un *Data Warehouse* sea posible de manipular, se construyen subconjuntos de información denominados *Data Marts*, que contienen datos personalizados y confeccionados para soportar los requerimientos analíticos de un determinado sector o función del negocio. El análisis de la información contenida en un *Data Mart* se obtiene con herramientas OLAP (*on-line analytical processing*) que involucran un tipo de procesamiento de datos que se caracteriza por permitir el análisis multidimensional dependiendo de la base de datos subyacente (Codd *et al.*, 1993) y en el caso particular de una base de datos relacional, se utiliza específicamente la herramientas OLAP denominada ROLAP (*relational on-line analytical processing*) que se describe como un cubo multidimensional. Según Kimball (2002) la palabra que mejor describe a ROLAP es la escalabilidad, dado que puede cubrir un amplio conjunto de datos, tanto datos precalculados y agregados, como las fuentes de datos que residen en la misma base de datos relacional. Los esquemas relacionales más utilizados para describir un *Data Mart* se denominan esquemas estrella, los cuales se caracterizan por poseer como componente principal; una Tabla de hechos conteniendo los datos de análisis y un conjunto de Tablas de dimensiones que describen las propiedades de los hechos. Ahora bien, previo a la obtención del *Data Mart* es necesario realizar un proceso ETL (*extract-transform-load*) que implica las tareas de capturar integrar, transformar, limpiar, reestructurar, validar, filtrar, analizar la

calidad y cargar datos en la base de datos relacional subyacente al *Data Mart* (Kimball *et al.*, 1998).

Por otro lado, dado que los datos de los censos se agrupan de acuerdo a la organización política del país, es factible que éstos puedan ser gestionados como una base de datos espacial. En tal caso, PostGIS, que es una extensión del sistema administrador de bases de datos PostgreSQL, permite gestionar objetos geográficos que cumplen con los estándares de interoperabilidad del Consorcio OpenGIS (OGC) (Martínez, 2012). Como sistema administrador de bases de datos espacial, posee la capacidad para representar y manipular datos geométricos, por tanto proporciona la tecnología para soportar aplicaciones, como sistemas de información geográficos (Guting y Schneider, 1995).

La solución desarrollada incorpora el uso de tecnología SOLAP (OLAP espacial), la que se refiere a un componente de software cuya arquitectura soporta bases de datos con propiedades espaciales (Rivest, Bérard *et al.*, 2005). Al ámbito al que se asocia el uso de una base de datos espacial, en el contexto de una solución de Inteligencia de Negocios, se le conoce como Inteligencia de Negocios Geoespacial (GeoBI). GeoBI involucra la integración de dos tecnologías, SIG e Inteligencia de Negocios (BI). Su uso conlleva un enorme potencial al combinar análisis espacial y visualización de mapas con herramientas y técnicas BI, tales como *data warehousing*, procesamiento analítico en línea (OLAP), herramientas de reportes, *dashboards* y minería de datos (Badard y Dubé, 2009).

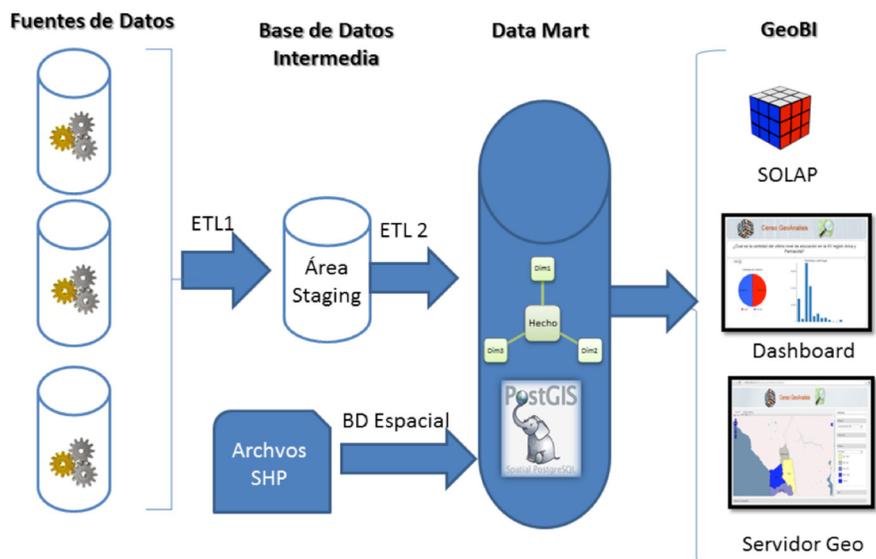


Figura 1. Arquitectura de una solución GeoBI.

En la Figura 1 se puede observar una arquitectura para una solución GeoBI, desde la recolección de las fuentes de información, pasando por un primer proceso ETL para ser cargados a una base de datos intermedia. Posteriormente, se procede un segundo ETL permitiendo la incorporación de información explícita de los datos y su inserción en un *Data Mart*. Por otra parte, se traspasan cartografías digitales a una base de datos espacial, permitiendo la integración de GIS en un *Data Mart* (Healey y Delve, 2007).

Un SIG es un conjunto de programas y aplicaciones informáticas que permiten la gestión de datos organizados en base de datos referenciados espacialmente, los que pueden ser visualizados mediante mapas (Moldes, 1995) que responden a dos formas de representación de datos: raster y vectorial. Raster representa un conjunto de unidades regulares constituidas por celdillas o píxeles en forma de mosaico. Vectorial es la estructura de datos que utiliza puntos, líneas y

polígonos para describir la información geográfica (Yañez Coto y Taboada, 2005).

Metodología

Para la construcción del *Data Mart* se realizó el proceso de análisis de la información del censo de los años 1992 y 2002 (entregados en formato SPSS por el INE), para obtener un esquema relacional que permitiese su gestión en un sistema de bases de datos (Cornejo et al., 2014). La Figura 2 muestra el modelo resultante de este análisis. La base de datos implementada a partir de este esquema se utilizó como un área de trabajo intermedia en el proceso de limpieza, transformación y carga (ETL). La etapa concluye con dos bases de datos intermedias: la base de datos del Censo 1992 y la del Censo 2002.

Posteriormente, se realizó el modelo e implementación del esquema estrella para soportar el *Data Mart* (Figura 3), que se estructura con cinco dimensiones: tres dimensiones representativas

de los módulos del Censo: vivienda, hogar y persona, una dimensión geográfica (región, provincia, etc.) y una dimensión del tiempo, que representa los años en que se han realizado cada Censo (1992 y 2002). Los datos propios de cada encuesta realizada se encuentran en la Tabla de hechos 'hecho_censo'.

En esta prueba inicial, se trabajó con un subconjunto de 18 preguntas del Censo: diez relacionadas con vivienda, seis con personas y dos con hogares. Cada atributo, exceptuando las claves primarias de las Tablas de dimensiones, está asociado a una pregunta de la encuesta Censo (Figura 3).

Un segundo proceso ETL involucró llevar los datos de las bases de datos intermedias de Censo al *Data Mart*. Este proceso permitió incorporar información explícita en este última, debido a que en los registros de los censos se mantenían datos codificados (se utilizaba el número de la opción seleccionada por el encuestado).

Para las consultas sobre respuestas a preguntas realizadas en ambos censos, se realizó un proceso de homogenización, considerando que en ciertos casos existían variaciones en la pregunta. Esas diferencias estaban asociadas al tipo de respuestas, a alternativas no comunes o al orden de las opciones permitidas. En la Tabla I se describen las opciones disponibles a una pregunta de censo de acuerdo a su formato 1992 y a su formato 2002. En aquellos casos en que no fue posible realizar esta homogenización, se generó grupo de respuestas por censo (año). Esto obligó a implementar diferentes vistas del *Data Mart* que dieran cuenta de estas situaciones.

En la Tabla I, se muestra el cuadro comparativo de opciones de respuesta a esta pregunta, lo que permitió su homogenización en el *Data Mart*. Para ello fue necesario llevar a cabo la limpieza (eliminar valores duplicados y corregir datos erróneos) y la transformación de los datos (cambios de formato, sustitución de códigos y generación de valores derivados, entre otros) con el propósito de realizar una conciliación entre ambos censos (coincidencia de códigos).

Por otro lado, para la visualización de la información de esta base de datos como un cubo multidimensional SOLAP se utilizó GeoMondrian, que es un servidor OLAP con soporte para operaciones de datos geospaciales (Badard y Dubé, 2009). La visualización de la información del cubo se realizó con la herramienta OLAP Jpivot (Dongen y Bouman, 2009), la cual interactúa con GeoMondrian para acceder a la base de datos del *Data Mart*.

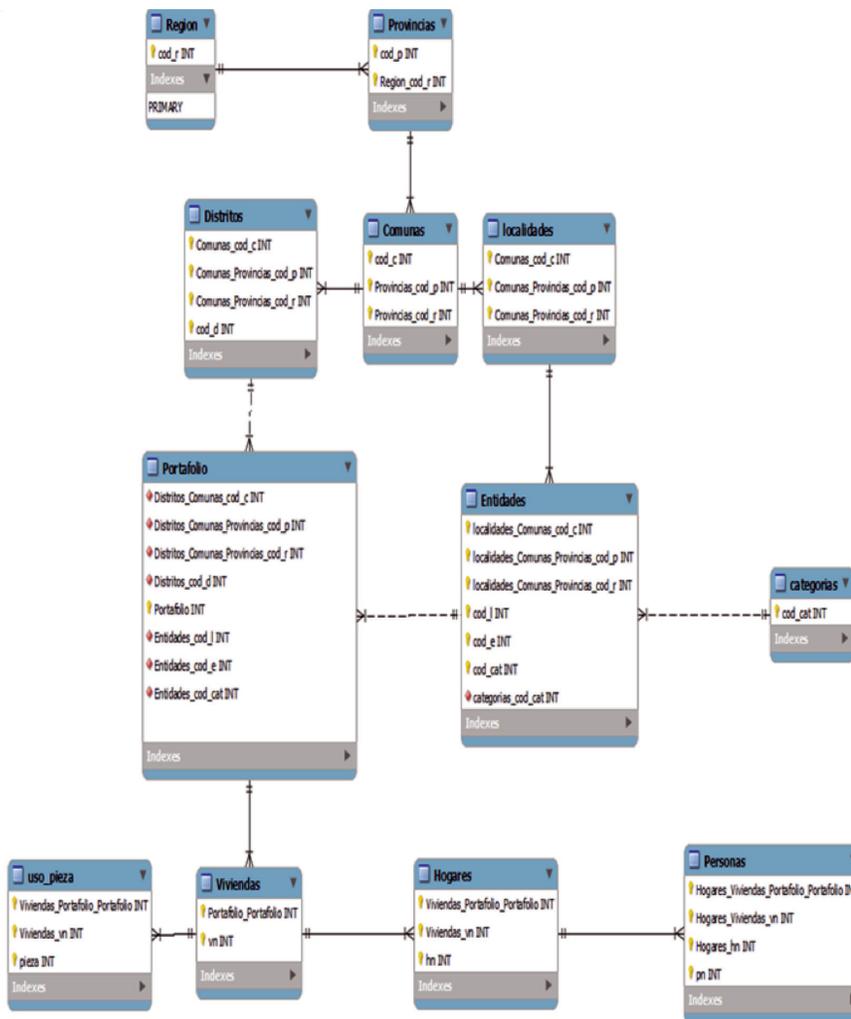


Figura 2. Modelo relacional de los censos.

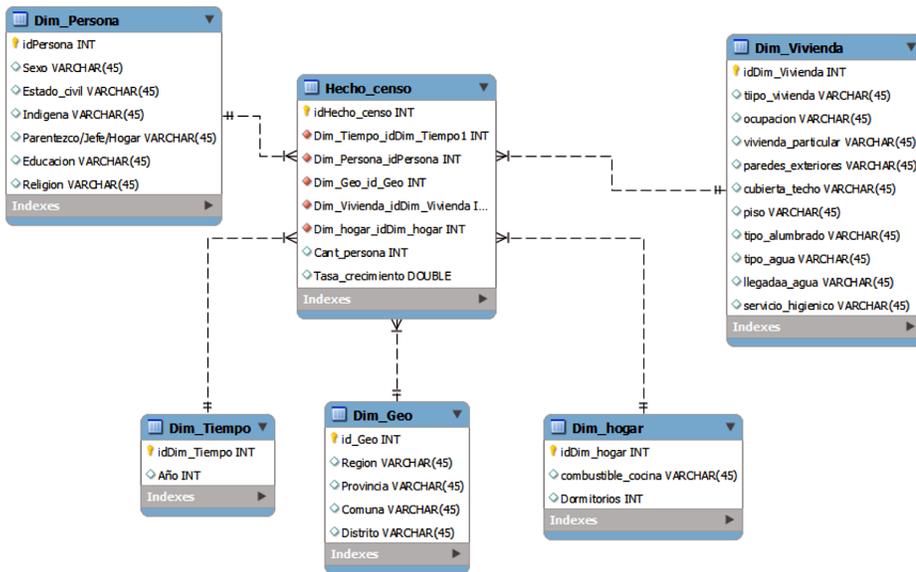


Figura 3. Modelo estrella del Data Mart.

TABLA I
CUADRO COMPARATIVO RESPUESTAS SEGUN TIPO
DE VIVIENDA EN CENSO '92 Y CENSO '02

Indique el tipo de vivienda:					
v1 1992		v1 2002		Homogenización de códigos	
				v1 1992	v1 2002
a	Vivienda particular	a	Vivienda Particular		
1	Casa	1	Casa	1	1
2	departamento en edificio	2	Departamento en edificio	2	2
3	Piezas en casa antigua o en un conventillo	3	Piezas en casa antigua o en un conventillo	3	3
4	Mejora, media agua	4	Mejora, media agua	4	4
5	Rancho, choza o ruca	5	Rancho choza		5
6	Móvil, carpa, vagón, etc.	6	Ruca		6
7	Otro	7	Móvil, carpa, vagón, etc.	5	7
		8	Otro	7	8
b	Vivienda Colectiva	b	Vivienda Colectiva		
8	Residencial, pensión	9	Residencial, pensión, hotel, hospital, motel, posada, etc.	8	9
9	Hotel, motel, posada			9	
10	Institución			10	
11	Otro			11	

Además, se implementó un conjunto de *dashboards* (Marcus, 2006) que entregan información gráfica en tiempo real combinando distintas variables por año. Para la elaboración de estos, se utilizó la herramienta *Community Dashboard Edition* (CDE; Alves *et al*, 2013). En la Figura 4 se muestra como ejemplo, la cantidad de hombres y mujeres, según último nivel de educación obtenido y relación con parentesco de hogar, durante el Censo 1992 (existiendo como opción el análisis de estos mismo datos con respecto al Censo de 2002).

Respecto de las herramientas para el GeoBI, se utilizó información geográfica de libre disposición en Chile (Albers, 2013), disponible en archivos vectoriales de formato Shape (ESRI, 1998). Para la creación de la base de

datos espacial en PostGIS a partir de los archivos Shape y, adicionalmente, reproyectar la cartografía para que estuviese de acuerdo a la escala WGS84 19S del

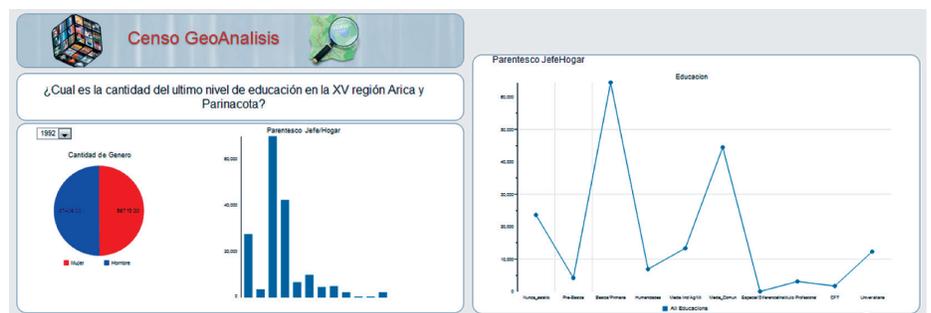


Figura 4. Dashboard, distribución poblacional por género, parentesco y nivel educacional.

servidor de mapas (Peña, 2006) se utilizó la herramienta Geokettle (Badard y Dubé, 2009). La base de datos espacial convive con el esquema estrella del *Data Mart*, permitiendo asociar la información analítica de éste con la información geográfica de la base de datos espacial.

Para la navegación de datos geoespaciales en el cubo, se implementó un servidor de mapas con capacidades SOLAP, que se denominó 'Geo Análisis Censo' (Figura 5). La arquitectura de este servidor incluye los siguientes componentes: 1) *openlayer*, librería responsable de insertar mapas dinámicos, para cualquier servidor de páginas web (OpenLayers, 2013); 2) extensiones DOJO para facilitar la implementación de paletas funcionales desarrolladas en Ajax; 3) librerías para la visualización de datos geoespaciales a través de consultas multidimensionales MDX (Badard y Dubé, 2009); y 4) JPivot, Geomondrian para visualizar y realizar consultas al cubo.

En la Figura 6 se aprecia el resultado de la visualización de la información del *Data Mart* en el servidor mapas de implementado. El mapa observado corresponde a la Región de Arica y Parinacota con sus divisiones comunales.

Resultados y Comparaciones

La aplicación de los procesos reseñados, en los datos de los censos 1992 y 2002 para Chile, dan como resultado la generación de una solución GeoBI que hemos llamado 'Geo Análisis Censo', a la que se puede acceder desde el servidor de mapas con capacidades SOLAP en la dirección http://146.83.109.233:8080/Plataforma_Censo/.

'Geo Análisis Censo' puede ser catalogada como una herramienta de Sistema de Información Geográfico de Participación Pública: PPGIS (McHugh *et al.*, 2008) para los datos Censo 1992 y Censo 2002. La Figura 6 muestra cada una de las funcionalidades de la herramienta donde destacan cinco puntos de interés:

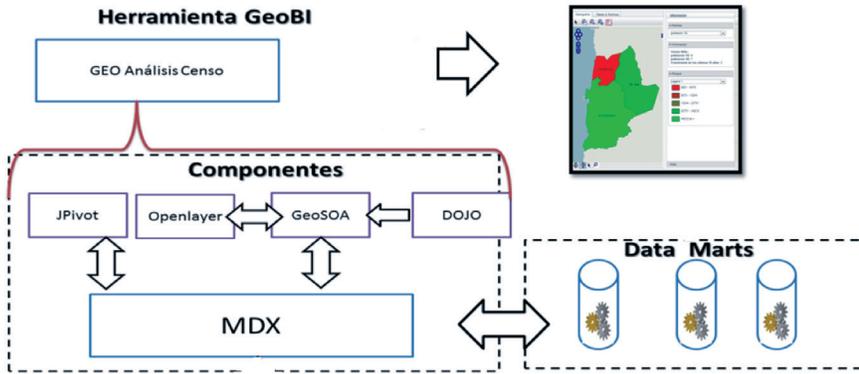


Figura 5. Arquitectura del servidor de mapas 'Geo Análisis Censo'.

más claro es el color de la zona menor es su valor de medición.

4. *Paleta de propiedades geográficas.* Permite desglosar la dimensión geográfica de su ámbito general a lo más específico. En este caso de provincia hasta distrito y viceversa.
5. *Frame Jpivot.* Permite realizar una conexión al servidor Geomondrian con el fin de realizar consultas multidimensionales.

En la Figura 7 se puede observar el Frame Jpivot, con el cual es posible realizar análisis de la información de censo. Éste permite seleccionar variables sobre una vista del *Data Mart*, para realizar consultas específicas.

El servidor 'Geo Análisis Censo' también admite como funcionalidad la posibilidad de realizar consultas multidimensionales en el servidor de mapas web. Esta funcionalidad se denomina 'paleta de filtro'. Ella incorpora un menú de dimensiones, donde es posible seleccionar datos pertenecientes a cada dimensión los que actúan como filtros (áreas destacadas en la Figura 8).

Adicionalmente, se implementó un conjunto de *dashboards*. Por ejemplo, uno de ellos describe la distribución de género de la población por región del país, el parentesco con respecto al jefe de hogar y el último nivel educacional cursado. Esta información se muestra para ambos censos en la Figura 4 mediante el uso de diferentes tipos de gráficos (torta, barra y línea).

Al seleccionar una de las áreas del gráfico de torta perteneciente al género se entrega información detallada del parentesco con respecto al jefe de hogar en un gráfico de barras. A su vez, al seleccionar un parentesco en el gráfico de barras se muestra una distribución del nivel de educación recibido en un gráfico de líneas.

Un segundo *dashboard* muestra la distribución de población a nivel de comuna de una región, clasificada por género y parentesco con respecto al jefe de hogar. Este *dashboard* incluye un componente de mapas, que permite ir seleccionando dinámicamente la comuna a analizar (Figura 9).

Finalmente, un tercer *dashboard* se refiere a consultas de población por comuna y distrito, mostrando distribución por relación de parentesco con el jefe de hogar y nivel educacional. Contiene un componente de mapas a nivel comunal, asociadas a una Tabla con datos del distrito de la comuna seleccionada (Figura 10). Cada componente tiene interactividad con el resto, desde el componente de mapas hasta el gráfico de torta que refleja el nivel de estudio de la población de acuerdo a los censos 1992 o 2002.

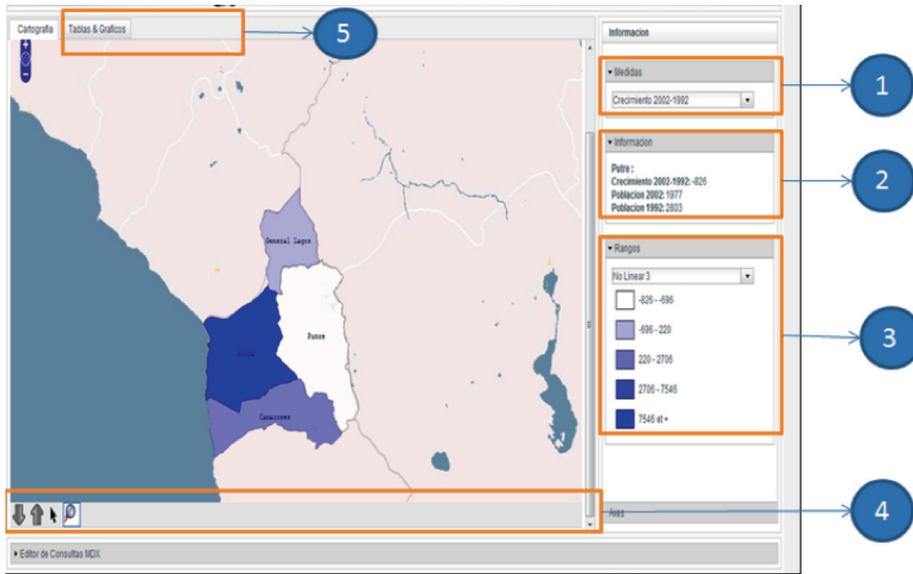


Figura 6. Geo Análisis Censo.

Zona	Categoría	Medidas		Poblacion 2002		Poblacion 1992	
		Crecimiento 2002-1992		Genero		Genero	
		Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Antofagasta	Conviviente/Pareja	2.869	3.421	4.967	38.077	2.098	34.656
	Hermano/a, Cuñado/a	426	24	2.806	2.313	2.380	2.289
	Hijo/Hijastro	4.517	4.067	50.436	48.875	45.919	44.808
	Jefe_Hogar	7.166	7.578	47.289	21.840	40.123	14.262
	Miembro_Hogar_Colectivo	8.943	252	14.389	1.882	5.446	1.630
	Nieto/a	1.467	1.421	9.469	8.856	8.002	7.433
	No_Pariente	535	297	4.107	2.752	3.573	2.455
	Otro_Pariente	4.095	2.143	7.837	5.808	3.742	3.665
	Padres/Suegros	65	415	783	3.116	718	2.701
	Servicio_Domestico_Puerta_Adentro	-7	-95	15	274	22	369
Calama	Transuntes	336	73	864	246	528	173
	Yerno/Nuera	356	219	2.469	1.961	2.113	1.742
	Conviviente/Pareja	1.411	-3.659	2.229	16.785	818	20.444
	Hermano/a, Cuñado/a	-122	-363	1.282	799	1.404	1.162
	Hijo/Hijastro	-3.170	-3.076	23.019	22.449	26.189	25.525
	Jefe_Hogar	-2.845	2.899	21.499	8.841	24.344	5.942
	Miembro_Hogar_Colectivo	1.390	149	3.018	581	1.628	432
	Nieto/a	572	372	3.310	3.072	2.738	2.700
	No_Pariente	433	138	1.682	922	1.249	784
	Otro_Pariente	1.188	462	2.989	2.022	1.801	1.560

Figura 7. Frame Jpivot.

1. *Paletas de medidas.* Puede escoger cualquier tipo de medida, cambiando el grado de color en el mapa de acuerdo a la medida escogida. Estas medidas son: cantidad poblacional del año 1992, cantidad poblacional del año 2002 y el crecimiento entre ambos años.
2. *Paletas información.* Muestra la información de las tres medidas de acuerdo a la región, provincia, comuna o distrito que se esté analizando.
3. *Paleta de rango de medida.* Muestra el rango de color de cada cantidad calculada en la región consultada (en este ejemplo por comuna). Mientras

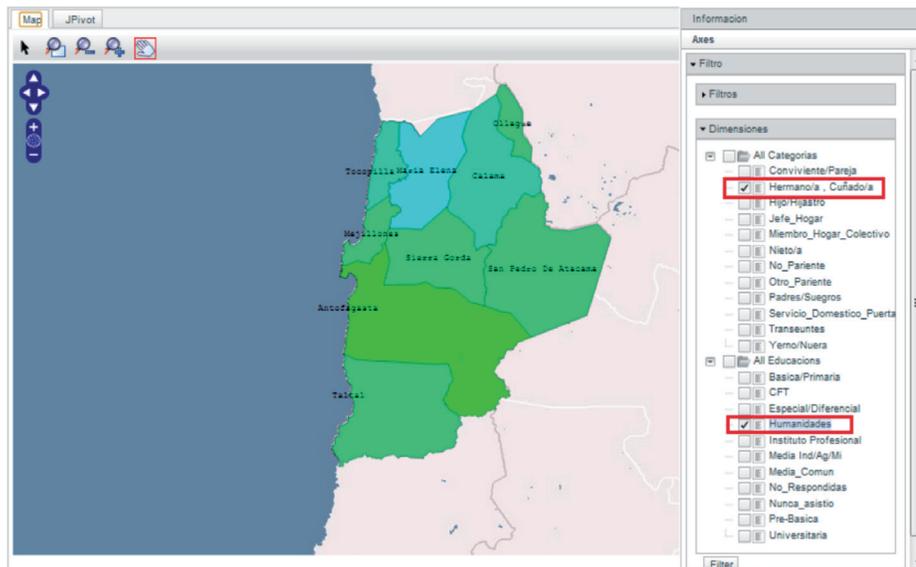


Figura 8. Paletas de filtro en 'Geo Análisis Censo'.

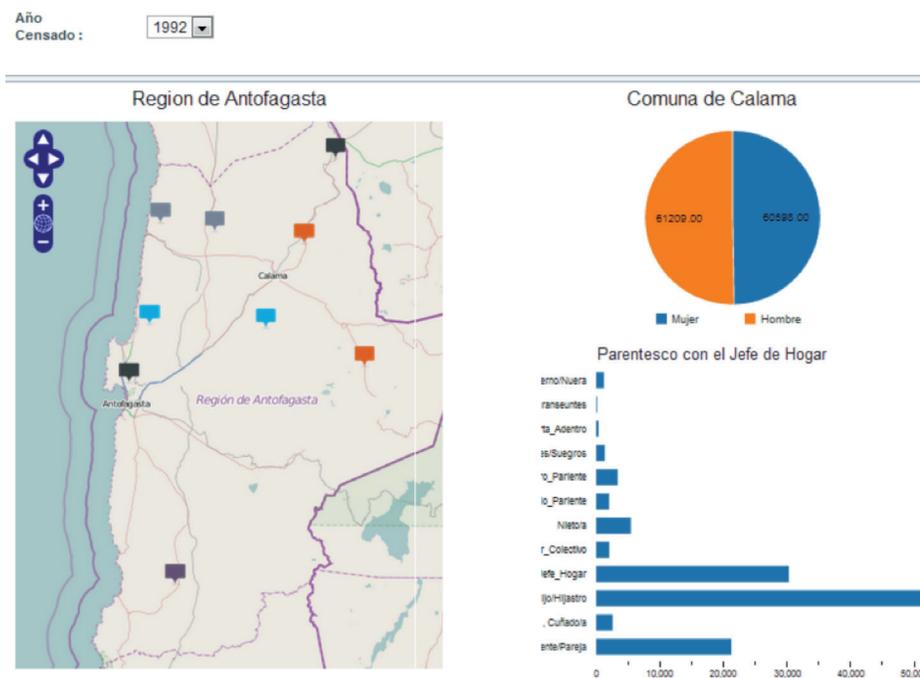


Figura 9. Dashboard, distribución poblacional a nivel de comuna por género y parentesco.

Análisis y Discusión

Análisis

Para validar la calidad de datos censales tratados se procedió a realizar un análisis comparativo entre la herramienta REDATAM+SP *Web Server* y la herramienta propuesta 'Geo Análisis Censo'. El análisis consistió en realizar consultas específicas del Censo en la II Región de Antofagasta, con la variable de parentesco con el jefe de hogar y el género del individuo perteneciente al módulo

persona. Al ingresar la consulta en 'Geo Análisis Censo' se muestra una cartografía de la comuna de Calama a nivel de distrito (Figura 11). Las zonas de color verde en el mapa reflejan la mayor cantidad de población en la comuna. La cartografía puede modificarse dinámicamente con los diferentes tipos de medidas que entrega esta herramienta.

REDATAM+SP *Web Server* entrega una lista donde se puede seleccionar sólo un tipo de variable, para la cual se muestra un mapa, una Tabla y un gráfico. Esta funcionalidad entrega

información estática, sin la capacidad de interactuar con el usuario y en cuanto al plano cartográfico, sólo realiza consultas a nivel de región. En el ámbito de Tablas y gráficos, REDATAM+SP permite realizar consultas con tres tipos de variables cruzadas, mientras que la solución 'Geo Análisis Censo' no solo permite realizar lo anterior sino que, además, entrega gráficos y Tablas dinámicas con las cuales se puede interactuar del modo que considere más conveniente.

En la Tabla II se muestra la información de la comuna de Calama a nivel de distrito, cruzando datos como, por ejemplo, el género del individuo con el jefe de hogar. Esta información se compara a través del tiempo, reflejando un claro impacto en el manejo de la información que entrega la herramienta propuesta.

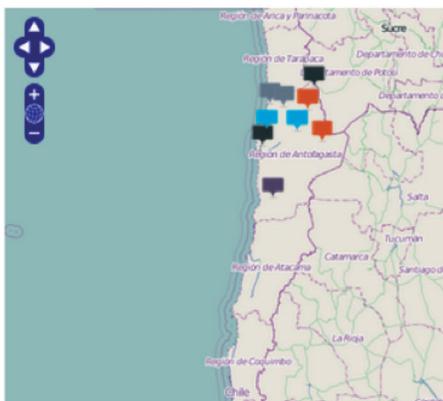
Por otra parte, las plataformas SIT y ICET (SIIGSA, 2014) se caracterizan por su sencillez y capacidad de interactuar con la cartografía e información asociada. Ambas son herramientas intuitivas y de fácil utilización para el usuario, características que comparten con 'Geo Análisis Censo'. La diferencia está en que las plataformas SIT e ICET no son catalogadas como soluciones GeoBI, pues no realizan cruzamientos de datos, obteniendo como resultado varias cartografías digitales para distintas consultas. La arquitectura SOLAP de 'Geo Análisis Censo' permite que los usuarios puedan realizar variados cruzamientos entre las variables existentes en el Censo (personas, viviendas o hogares), dando la posibilidad de construir consultas acordes a sus necesidades.

El uso de instrumentos cuantitativos, como los censos, puede ser utilizado para validar estudios (Aravena y Álvarez, 2012). Adicionalmente, el uso complementario de tecnologías SIG puede indicar las zonas en las que se deben enfocar un análisis en particular (Castro *et al.*, 2004).

Discusión

REDATAM+SP, pese a su uso extensivo, posee funcionalidades básicas SIG y no realiza comparaciones temporales entre censos de diferentes años. Su versión *online* solo tiene una desagregación espacial a nivel de comuna en el caso de Chile, con la antigua división política de 13 regiones. Mientras que, en esta propuesta, la solución ofrece una desagregación espacial a nivel de distrito y respeta la división política actual de 15 regiones. Uno de los aspectos más relevantes consiste en la capacidad de realizar comparaciones de datos entre los diferentes años en que fue aplicado el censo (1992 y 2002).

Mapa II Region de Antofagasta data de 1992

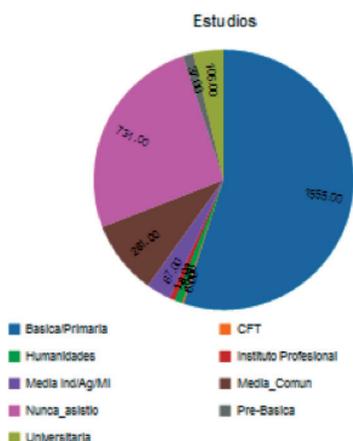


Comuna de San Pedro de Atacama

Search:		
Distrito	Mujer	Hombre
Total	1356	1473
San Pedro de Atacama	749	820
Socaire	264	307
Toconao	343	346

Showing 1 to 4 of 4 entries

Comuna de San Pedro de Atacama



Search:			
Categoría	Mujer	Hombre	Total
Total	721	824	1545
Conviviente/Pareja	174	17	191
Hermano/a , Cuñado/a	19	26	45
Hijo/Hijastro	277	279	556
Jefe_Hogar	147	327	474
Miembro_Hogar_Colectivo	40	75	115
Nieto/a	28	51	79
No_Pariente	8	23	31
Otro_Pariente	11	12	23
Padres/Suegros	11	1	12

Showing 1 to 10 of 13 entries

Universitaria

Figura 10. DASHBOARD, distribución poblacional por comuna y distrito, incluyendo parentesco y nivel educacional.

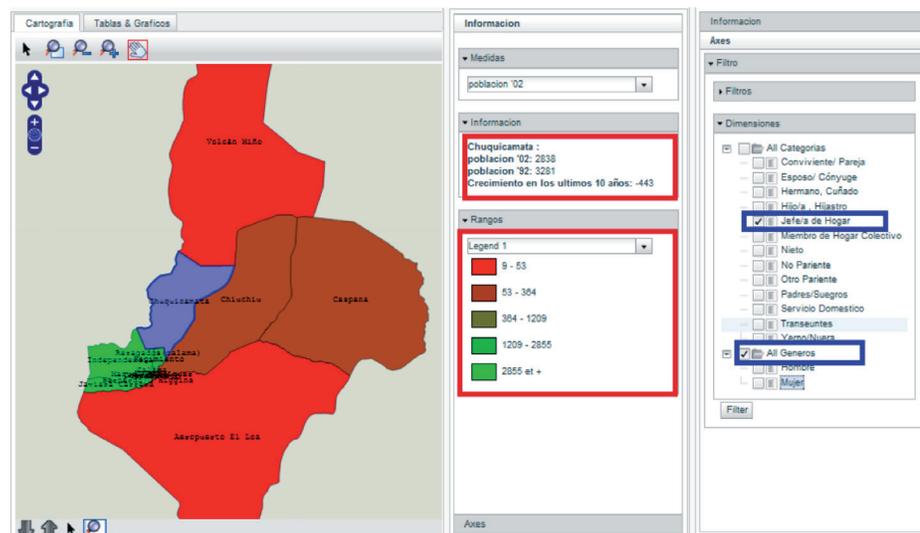


Figura 11. Cartografía de 'Geo Análisis Censo'.

REDATAM+SP Web Server entrega información cartográfica, pero en archivos raster, sin posibilidades de interactividad. En cambio, la solución propuesta

en este trabajo entrega información cartográfica dinámica, permitiendo interactuar con el servidor de mapas y, a diferencia de las plataformas SIT e ICET, puede realizar

cruces de distintas variables en una misma cartografía, para distintos niveles territoriales (provincia, comuna o distrito).

Este trabajo contribuye a la integración de tecnología SIG e Inteligencia de Negocios, para tratar datos del Censo en Chile. Esto permite desarrollar un repositorio de información temporal de datos de censos relacionada con información geoespacial. Sin embargo, todavía queda la tarea de incorporar mapas políticos de Chile anteriores al año 2007. Esto implica que, al realizar una desagregación espacial a nivel de distrito, es posible ver, en algunos casos, una cartografía sin coloración e información, debido a que estos no existían en la época del Censo 1992. La versión actual de 'Geo Análisis Censo' entrega información de tres regiones: la XV Región de Arica y Parinacota, la I Región de Tarapacá y la II Región de Antofagasta. A futuro se plantea incorporar la información censal para el resto de las regiones de Chile.

Conclusiones

El Censo es considerado como uno de los instrumentos de medición más importantes del sistema estadístico nacional (Chackiel, 2002). A partir del año 2000 se ha caracterizado por incorporar importantes cambios en el aspecto tecnológico, los que inciden en la mejora de calidad de los datos y en la difusión de información. Una de las herramientas responsables en la difusión de información censal es REDATAM (Jaspers y Poulard, 2001), utilizado hoy en día a nivel internacional. La motivación de este trabajo se centró en la aplicación de tecnología SIG e Inteligencia de Negocios de libre disposición en el proceso de visualización y análisis de los datos. Con la tecnología GeoBI fue posible realizar análisis de los datos Censo 1992 y Censo 2002, permitiendo visualizar los datos con el apoyo de mapas cartográficos. Al incorporar procesamiento SOLAP se pudo obtener una herramienta de acceso ágil y directo a base de datos espaciales, ofreciendo un nivel de información detallada sincronizada a través de mapas y Tablas (McHugh *et al.*, 2008). La solución 'Geo Análisis Censo' permite una exploración intuitiva sobre una base de datos espacial, presentando una visualización y manipulación interactiva de componentes que sincronizan mapas y Tablas. Además, con el mecanismo OLAP, es posible proporcionar una capacidad analítica rica y variada que va más allá de todo lo visto anteriormente en términos de análisis temporal de censos (Healey y Delver, 2007). Este tipo de solución con tecnología GeoBI ofrece una herramienta de fácil acceso para los

COMPARACIÓN DEL CENSO CON EL JEFE DE HOGAR EN LA COMUNA DE CALAMA, II REGIÓN DE ANTOFAGASTA, SEGÚN CENSO '02 Y CENSO '92

		Población '02	Población '92	Variación intercensal	
Comuna	Calama	Total Hombre	35.937 25.563	30.286 24.344	5.651 1.219
Nombre Distinto	Aeropuerto El Loa	Total	52	0	52
		Hombre	49	0	49
	Bernardo O'Higgins	Total	5.233	0	5.233
		Hombre	3.782	0	3.782
	Calama	Total	3.885	26.998	-23.113
		Hombre	2.897	21.370	-18.473
	Caspana	Total	124	0	124
		Hombre	82	0	82
	Cementerio	Total	1.122	0	1.122
		Hombre	857	0	857
	Chacabuco	Total	4.974	0	4.974
		Hombre	3.357	0	3.357
	Chiuchiu	Total	151	0	151
		Hombre	104	0	104
	Chuquicamata	Total	2.838	3.281	-443
		Hombre	2.422	2.968	-546
	Independencia	Total	4.164	0	4.164
		Hombre	2.786	0	2.786
Javiera Carrera	Total	2.972	0	2.972	
	Hombre	2.030	0	2.030	
Maipú	Total	5.568	0	5.568	
	Hombre	3.572	0	3.572	
Manuel Rodríguez	Total	1.948	0	1.948	
	Hombre	1.377	0	1.377	
Regimiento	Total	2.533	0	2.533	
	Hombre	1.959	0	1.959	
Rezagados	Total	364	0	364	
	Hombre	282	0	282	
Volcán Miño	Total	9	7	2	
	Hombre	7	6	1	

usuarios que no poseen conocimiento experto en tecnología. Cabe destacar que 'Geo Análisis Censo' ha sido probada en conjunto con REDATAM para confirmar la calidad de la información obtenida. Las funcionalidades que se incorporaron para mapas interactivos (SIG) y el repositorio histórico de datos Censos (Inteligencia de Negocio), proyectan un punto de vista diferente para realizar soluciones de este tipo de instrumento de medición.

REFERENCIAS

Albers C (2013). *Coberturas SIG para la Enseñanza de la Geografía en Chile*. www.rulamahue.cl/mapoteca/catalogos/chile.html (Cons. 13/12/2013).

Alves P, Kemenade R, Phillips A (2013) *CDE and the Technology Underneath it (CDF, CDA and CCC) Allows the Development and Deployment of Pentaho Dashboards in a Fast and Effective Way*. Obtenido de <http://www.webdetails.pt/ctools/cde.html> (Cons. 13/12/2013).

Aravena A, Álvarez C (2012) Juventud, migración y discriminación en el Chile contemporáneo. *Ultima Década N° 36*: 127-140.

Badard T, Dubé E (2009) *Enabling Geospatial Business Intelligence*. http://geosoa.seg.ulaval.ca/~badard/article-tbadard-osbr_2009-long_version-enabling_geospatial_bi.pdf (Cons. 13/06/2013).

Baker M (2013). *Publishing an Analysis Schema Using Schema Workbench*. <http://wiki.pentaho.com/display/analysis/Publishing+an+Analysis+Schema+Using+Schema+Workbench> (Cons. 13/12/2013)

Castro V, Varela V, Aldunate C, Araneda E (2004) Principios orientadores y metodología para el estudio del Qhapaqñan en Atacama: Desde El Portezuelo del Inka hasta Río Grande. *Rev. Antropol. Chil.* 36: 439-451.

CELADE (2004) *Familia REDATAM*. www.eclac.cl/cgi-bin/getProd.asp?xml=/redatam/noticias/software/5/7395/P7395.xml&xml=/redatam/tpl/p17f.xslybase=/redatam/tpl/top-bottom.xsl (Cons. 26/12/2013).

CELADE (2013) *Bienvenido a la Ayuda REDATAM+SP Process*. www.ec2013lac.cl/redatam/publicaciones/sinsigla/xml/2/10622/Man-RedatamSPPEsp1.pdf (Cons. 26/12/2013).

CEPAL (2013) *Antecedentes*. www.eclac.cl/celade/Antecedentes50.htm (Cons. 20/12/2013).

Chackiel J (2002) *Los Censos en América Latina: Nuevos Enfoques*. <http://www.cepal.org/publicaciones/xml/3/13143/2.pdf> (Cons. 01/08/2013).

Codd E, Codd S, Salley C (1993) *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate*. Codd y Date. Inc. Technical Report. 26 pp.

Conesa J, Curto J (2010) *Introducción a Bussines Intelligence*. UOC. Barcelona, España. 238 pp.

Cornejo R, Navarrete M, Aroca P, Valdivia R, Aracena S (2012) *Desarrollo de una Base de Datos Integrada de Censo y Encuesta. Mediante el Uso de Elementos de Inteligencia de Negocios y SIG*. *INGENIARE 22*: 205-217.

Dongen R, Bouman J (2009) *Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL*. Wiley. Indianapolis, IN, EEUU. 604 pp.

ESRI (1998). *ESRI Shapefile Technical Description*. www.esri.com/library/whitepapers/pdfs/shapefile.pdf (Cons. 01/10/2013).

Guting M, Schneider (1995) Realm-Based Spatial Data Types: The ROSE Algebra. *VLDB Journal 4*: 243-286.

Healey R, Delve J (2007) Integrating GIS and data warehousing in a Web environment: A case study of the US 1880 Census. *Int. J. Geogr. Inf. Sci.* 21: 603-624.

INE (2002) *Censo 2002 Resultados. Vol. I. Población País - Región*. Instituto Nacional de Estadística. Chile.

INE (2002) *Manual Censal 2002*. Instituto Nacional de Estadística. Chile.

Jaspers D, Poulard S (2001) *El Software Redatam Para Divulgación y Análisis De Datos Censales*. 20ª Conf. sobre Censos de Población. 19-21/06/2001. Ulán Bator, Mongolia. pp. 321-341.

Kimball R (1992) *The Data Warehouse Toolkit*. Wiley Computer Publishing. New York. 464 pp.

Kimball R, Ross (2002) *The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling*. 2ª ed. Wiley. Nueva York, EEUU. 440 pp.

Kimball R, Reeves L, Ross M, Thornthwaite W (1998) *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. Wiley. Nueva York, EEUU. 800 pp.

Marcus A (2006) Dashboards in your future. *Art Prototyp.* 13: 48-60.

Martínez JC (2012) *PostGIS 2. Análisis Espacial Avanzado*. CreateSpace. España. 500 pp.

McHugh R, Roche S, Bédard Y (2008) Towards a SOLAP Based public participation. *J. Environ. Manag.* 90: 2041-2054.

Moldes FJ (1995) *Tecnología de los Sistemas de Información Geográfica*. RA-MA. Madrid. 190 pp.

ODEPA (2014) *ICET Sistema de Consultas Estadístico Territorial*. <http://icet.odepa.cl/> (Cons. 01/07/2014).

OpenLayers (2013). *OpenLayers: Free Maps for the Web*. <http://openlayers.org/> (Cons. 15/12/2013).

Peña J (2006) *Sistemas de Información Geográfica Aplicados a la Gestión del Territorio*. Club Universitario. España. 310 pp.

Rivest S, Bédard Y, Proulx M, Nadeu M, Hubert F, Pastor J (2005). SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal

- exploration and analysis of data. *J. Photogram. Rem. Sens.* 60: 17-33.
- Roldán RM (2013) *Pentaho Data Integration (Kettle) Tutorial*. Obtenido de [http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integrati+on+\(Kettle\)+Tutorial](http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integrati+on+(Kettle)+Tutorial) (Cons. 13/11/2013).
- SIIGSA (2014). *Servicio Integrados en Información Geográfica S.A.* www.sii-gsa.cl/ (Cons. 21/07/2014).
- SIT (2014) *Sistema de Información Territorial Social*. <http://sit.ministeriodesarrollosocial.gob.cl/> (Cons. 21/06/2014).
- Wickramasuriya R, Ma J, Berryman M, Perez P (2013) Using geospatial business intelligence to support regional infrastructure governance. *Knowl.-Based Syst* 53: 80-89.
- Yañez Coto J, Taboada J (2005). *Sistemas de Información Medioambiental*. Netbiblo. España. 272 pp.

DEVELOPMENT OF A GEOBI SOLUTION FOR THE OBSERVATION AND ANALYSIS OF CENSUS INFORMATION IN CHILE

Sebastián Aracena Aguirre, Ricardo Valdivia Pinto, Mónica Navarrete Álvarez and Robert Cornejo Yañez

SUMMARY

A Business Intelligence solution for the 1992 and 2002 censuses in Chile is described, aiming to integrate information from both periods so as to facilitate the observation and time-space data analysis. The solution involves the implementation of a Data Mart that includes geospatial information (GeoBI) and allows using GIS technology and spatial online analytical processing (SOLAP) tools for displaying information. To extend the dissemination of this information, a web map server 'Geo Analysis Census' has been implemented, through which it is possible

to dynamically interact with the population data, considering the use of cubes, maps and dashboards. The existing alternative in the market is the REDATAM tool, widely used in Latin America to manage information census. However, it does not permit integrated queries between censuses in different years, nor to interact with cartographical data visualization. Therefore, this work contributes to the observation and census geoinformation analysis by using open access tools that enable a more flexible and powerful integrated analysis.

DESENVOLVIMENTO DE UMA SOLUÇÃO GEOBI PARA A OBSERVAÇÃO E ANÁLISE DA INFORMAÇÃO DE CENSO NO CHILE

Sebastián Aracena Aguirre, Ricardo Valdivia Pinto, Mónica Navarrete Álvarez e Robert Cornejo Yañez

RESUMO

Descreve-se uma solução de Inteligência de Negócios para os censos realizados no Chile nos anos 1992 e 2002, com o objetivo de integrar a informação de ambos os períodos para facilitar a observação e análise espaço-temporal dos dados. A solução envolveu a implementação de um Data Mart que inclui informação geoespacial (GeoBI), o que permite o uso de tecnologia SIG e ferramentas espaciais de processamento analítico em linha (SOLAP) para a visualização da informação. Para ampliar o âmbito de difusão de esta informação, tem sido implementado um servidor de mapas web 'Geo Análise Censo',

através do qual é possível interagir dinamicamente com a informação de população, considerando o uso de cubos, cartografias e dashboards. A alternativa existente no mercado é REDATAM, ferramenta amplamente utilizada na América Latina para gerenciar a informação de censos. No entanto, esta não permite realizar consultas integradas entre censos de diferentes anos, nem interagir com a visualização cartográfica dos dados. Por tanto, este trabalho contribui para a observação e análise da geoinformação censal com ferramentas de livre acesso que possibilitam uma análise integrada mais flexível e potente.