

Simonazzi Domínguez, W., Ramos Jordan, M.J. y Ubach, R. (2010): Integración y explotación de datos ambientales y socioeconómicos a escala europea. En: Ojeda, J., Pita, M.F. y Vallejo, I. (Eds.), *Tecnologías de la Información Geográfica: La Información Geográfica al servicio de los ciudadanos*. Secretariado de Publicaciones de la Universidad de Sevilla. Sevilla. Pp. 1.385-1.392. ISBN: 978-84-472-1294-1

INTEGRACIÓN Y EXPLOTACIÓN DE DATOS AMBIENTALES Y SOCIOECONÓMICOS A ESCALA EUROPEA

Simonazzi Dominguez, Walter¹, Ramos Jordan , María Jose¹, Ubach, Raquel¹

(1) Centro Temático Europeo de Usos del Suelo e Información Espacial (ETC-LUSI). Universidad Autónoma de Barcelona, Facultad de Ciencias, Torre C5-S, 4ª planta, walter.simonazzi@uab.es, mariajose.ramos@uab.cat, ubach.r@gmail.com

RESUMEN

Dentro del proyecto Land and Ecosystem Accounting (LEAC) que se realiza en el contexto de trabajo de la Agencia Europea de Medio Ambiente, se ha desarrollado una metodología y posterior implementación de una herramienta que permite analizar los usos del suelo, sus cambios y tendencias para diferentes años. Además, esta herramienta permite la integración no sólo de datos medioambientales sino también la incorporación de otro tipo de datos como son los socio-económicos a través de diferentes metodologías; como por ejemplo: el Producto Interior Bruto, la población activa o zonas de influencia urbana. Esta integración de datos de diferentes temáticas permite realizar un análisis de la información disponible desde una perspectiva mucho más holística.

En el presente artículo se describe la naturaleza de los datos, así como los distintos procesos que se llevan a cabo con los datos espaciales y temáticos mediante el uso de diferentes programas de Sistemas de Información Geográfica (SIG), su introducción en una base de datos relacional y finalmente, mediante el uso de la tecnología On-line Analytical Processing (OLAP), la generación de una pequeña base de datos, denominada "cubo" en terminología OLAP, que contiene todas las combinaciones y posibles agregaciones de la información de partida.

El usuario final es capaz de realizar una gran multitud de consultas espaciales de forma inmediata, sin la necesidad de utilizar ningún programa SIG utilizando para ello simples paquetes de ofimática, como por ejemplo Excel que ofrece la posibilidad de generar tablas y gráficos. Sin embargo, el futuro de dicha herramienta pasa por el desarrollo de una plataforma Web que permita la explotación y análisis de estos cubos en forma de gráficos, tablas y mapas interactivos. Es por este motivo que en el contexto del proyecto LEAC se desarrollará un prototipo de herramienta Web, que facilite el uso y acceso universal a la información con la funcionalidad anteriormente descrita.

Palabras Clave: OLAP, Integración de datos, Sistemas de Información Geográfica, Datos ambientales y socioeconómicos.

ABSTRACT

Within the Land and Ecosystem Accounting (LEAC) project developed in the context of work of the European Agency of Environment, it has been developed a methodology, which was then implemented into a tool that allows to analyze the uses of the land, its changes and tendencies for different years in Europe. In addition, this tool allows the integration not only of environmental data but also the incorporation of others types of data, as for example the socio-economic information, through different methodologies. Example of variables could be the Gross Domestic Product, active population or zones of urban influence. This integration of different thematic data allows undertaking an analysis of the available information available from a holistic perspective. In the present article the nature of the data is described, as well as the different processes to involve spatial and thematic data by means of the use of

different Geographic Information System (GIS) tools ,its introduction in a relational database and finally, by means of the use of the technology On-line Analytical Processing (OLAP), the generation of a small data base, denominated “cube” using the OLAP terminology which contains all the possible combinations and aggregations of the information included in the cube.

In this way, the end user is able to analyze of space consultations of immediate form, without the necessity to use any GIS program, instead of that using common software packages, like a spreadsheet, Excel in example, offers the users the possibility of generating tables and graphs. Nevertheless, the future of this tool goes through the development of a Spatial Web platform which will allows in the future the analysis of the cubes generating graphs, interactive tables and coropleth maps. This is the reason why in the context of the LEAC project a prototype of this tool has been planned to be developed, tool that will facilitate the exploratory analysis of the results through maps, charts and tables and a will also bring an universal access to the information and functionality due to it will be based in the Web2.0 technology.

Key words: Technology OLAP, Integration of data, GIS, environmental and socioeconomic Data

INTRODUCCIÓN

En el año 2001, momento en el cual el Centro Temático Europeo en Medio Ambiente Terrestre comenzó a desarrollar su actividad, la oficina europea de estadística, Eurostat, propuso testear la metodología LEAC (Land and Ecosystems Accounts) con vistas a su posible aplicación a los datos generados en el proyecto CORINE Land Cover (CLC). Dos casos de estudio fueron desarrollados y fueron seleccionados gracias a la disponibilidad de datos de cambios en usos del suelo.

- La costa europea (proyecto LACOST, para los años 1975 y para el año 1990)
- Cuatro países de Europa Central (República Checa, Eslovaquia, Hungría y Rumania), que habían hecho una evaluación de la Corine Land Cover para 1975.

Durante los años 2004 y 2005, los productos del proyecto Corine Land Cover generados a nivel Europeo se pusieron a disposición de forma gradual. Esta fue la primera vez que la mayor parte de la extensión de Europa quedó mapeada para dos fechas 1990 y 2000, hecho que permitió derivar una capa de cambios de usos del suelo.

La metodología LEAC, testada en los dos estudios piloto anteriormente mencionados, se adaptó a los nuevos conjuntos de datos, hecho que suponía la posibilidad de aplicar la metodología a Corine Land Cover 1990 y 2000, así como las correspondientes capas de cambios generadas.

Actualmente, y con la finalización de CLC para el año 2006, se nos presenta la posibilidad de aplicar la metodología LEAC para las tres fechas disponibles en CLC (1990, 2000 y 2006) y de esta forma analizar cambios y tendencias en los usos del suelo en Europa .

El objetivo principal de la metodología LEAC es proporcionar un acceso fácil y comprensivo a los datos generados en el proyecto Corine desde una perspectiva analítica; permitiendo cuantificar el “stock” y los “cambios” en cada clase de CLC para una determinada unidad de análisis, así como la cuantificación de los cambios en la cobertura del suelo producidos entre dos fechas.

La metodología LEAC también provee de una clasificación en “flujos” de los cambios en la cobertura del suelo que suponen una forma más comprensiva de analizar los cambios.

Los beneficios que aporta la metodología LEAC en comparación al análisis convencional que se puede hacer con un software SIG, es que contempla el procesado de toda la información a nivel Europeo y de cuantas unidades de análisis se requieran en un solo proceso continuo, generando una base de datos en las que todas las posibilidades de consulta se precaculan, permitiendo que posteriormente, las estadísticas basadas en CLC puedan ser consultadas de una forma fácil y rápida por el usuario final.

METODOLOGÍA Y DATOS QUE INTERVIENEN EN LEAC

La metodología LEAC se basaba en el análisis de los “stocks” y cambios para dos fechas de CLC, 1990 y 2000. Durante 2010 se espera que se complete CLC para el año 2006, lo que supondrá la posibilidad de disponer de una

tercera fecha en CLC y así poder aplicar la metodología para las fechas 2000 y 2006 permitiéndose el análisis de las tendencias en la cobertura del suelo en Europa.

En cualquier caso, y tomando como ejemplo las fechas 1990 y 2000, los productos finales del proyecto CLC 2000 se toman como principal insumo para la construcción de la bases de datos LEAC. Los productos del proyecto Corine son entregados por cada país que interviene en el proyecto al Centro Temático Europeo de Usos del Suelo e Información Espacial, en el cual se aplica un control de calidad de los datos y se procede a la generación de una capa continua para Europa. El producto final consiste en tres capas geográficas:

- CORINE Land Cover 2000
- CORINE Land Cover 1990 revisado.
- Los cambios que se han producido entre el año 1990 y 2000.

Estas capas proporcionan información sobre la superficie ocupada por diferentes tipos de cubierta del suelo y los cambios ocurridos en estas superficies entre dos fechas (1990 - 2000, 2000 - 2006) permitiéndonos analizar la información contenida por país como unidad de análisis territorial.

Dado que el objetivo de LEAC es permitir al usuario el análisis a diferentes unidades de análisis tales como:

- A nivel administrativo: límites a diversas escalas armonizados a nivel Europeo: NUTSO, NUTS1, NUTS2 y NUTS3, etc.
- Unidades físicas, tales como las cuencas, cuencas marinas, regiones biogeográficas, etc.

El primer objetivo en la metodología LEAC es asignar estas unidades territoriales a los datos *Corine Land Cover*. Para ello la metodología utilizada pasa por la integración de los diferentes tipos de información espacial a una malla de referencia de 1km² (*1km² European reference grid*) por medio de la intersección de cada unidad territorial con la malla de referencia. El uso de un sistema de referencia estándar como unidad territorial mínima facilita la posterior integración de otras fuentes de datos, tales como datos socio-económicos pudiéndose de esta forma su posterior análisis de forma conjunta con los datos de los usos del suelo, lo que a su vez permite establecer correlaciones entre los usos y cambios en los usos del suelo y su impacto en las variables socio-económicas.

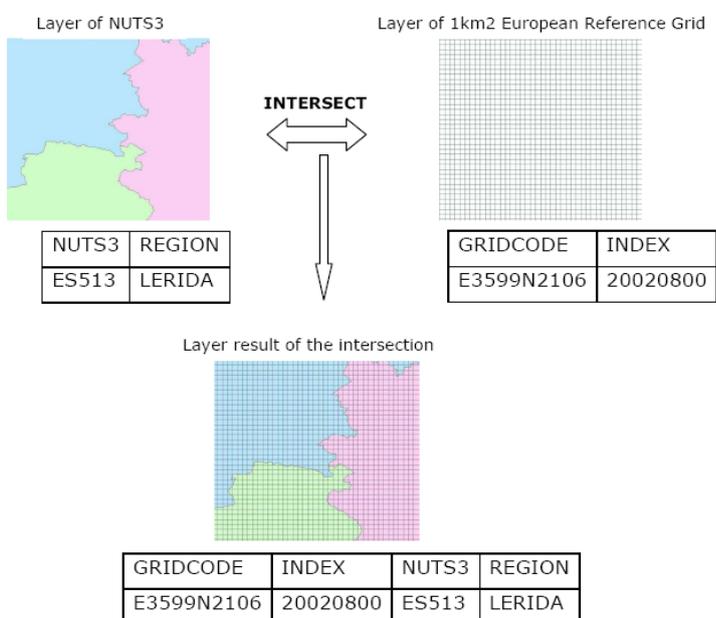


Figura 1. Intersección de Unidades administrativas en la malla de referencia.

INTEGRACIÓN DE DATOS DE COBERTURA DEL SUELO

El paso previo a la integración en la malla de referencia de todos los datos que intervienen en LEAC, es la preparación de la base de datos de cambios. De cara a obtener una base de datos coherente que nos permita identificar los cambios entre dos fechas de CLC y conocer no solo donde se ha producido un cambio en la cobertura del suelo sino que tipo el cambio se ha producido, se hace necesario construir un código único de cambio por medio del uso de los códigos asignados a cada clase de CLC. Para ello, y en las áreas donde se ha producido un cambio en la cobertura del suelo se combinan los códigos asignados para cada clase de CLC de ambas fechas, de tal manera que se puede generar un código de cambio que nos permitirá identificar en un área determinada si se ha producido un cambio en la cobertura del suelo así como identificar el cambio en si.

Para ilustrar lo descrito anterior consideremos que en un área determinada se ha producido un cambio entre los años 1990 y 2000 de “coniferous forest” que según la nomenclatura de CLC esta codificado como 311 a “continuous urban fabric” codificado como 111. Según el ejemplo anterior, el código de cambio resultante será 311111. Continuando con este ejemplo, y en caso de que no existiera un cambio en la cobertura del suelo, el código resultante sería 311311.

Los posibles cambios entre las clases de CLC están recogidos en lo que se ha denominado “Land Cover flows” que se comentarán más adelante.

Dependiendo del formato de los datos (raster o vector) se pueden utilizar diferentes procesos SIG para obtener el mismo resultado: Una tabla que contenga la superficie total para cada tipo de cambio en cada celda de la malla de referencia.

Esta tabla se considera la tabla principal del modelo de datos de LEAC. La tabla “LEAC_DATA” contendrá tres campos con la siguiente información:

- Código de celda de la malla de referencia.
- Código de Cambio (más no cambios)
- Superficie (en hectáreas)

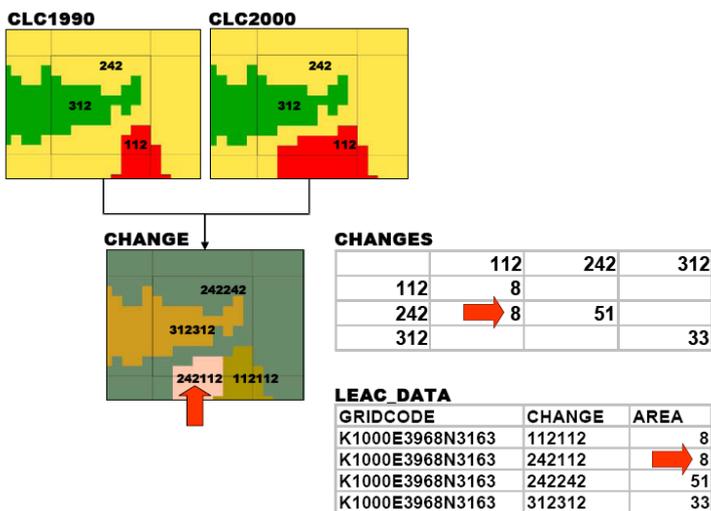


Figura 2. Creación de la tabla LEAC_DATA

INTEGRACIÓN DE LAS UNIDADES TERRITORIALES

Una vez que los datos de CLC se han asignado a las celdas de la malla de referencia se hace necesario asignar a cada celda las distintas unidades territoriales. Estas unidades territoriales se utilizarán para agrupar las celdas en el momento de analizar los cambios o el "stock" en la cobertura del suelo. Además, estas unidades territoriales serán utilizadas para reportar los resultados de los resultados obtenidos tras la consulta a la base de datos LEAC.

En el contexto de la LEAC nos referimos a estas unidades territoriales como LARUS (*Land Analytical and Reporting Units*). Cada celda contiene un valor único para un LARU dado. Para asignar un valor de LARU a una celda de la malla de referencia, utilizamos el criterio de "área máxima", que se explicará en el siguiente capítulo.

Otro factor a tener en cuenta es que debido a las diferencias geométricas entre las capas de CLC y las unidades territoriales (LARUS), algunas de las celdas de la malla de referencia cercanas a la costa no heredan datos. Para minimizar este error se aplica un proceso SIG llamado "proximity" por el cual los atributos de los LARUS se interpolan a las zonas donde anteriormente no existían datos debidos a diferencias en las geométricas.

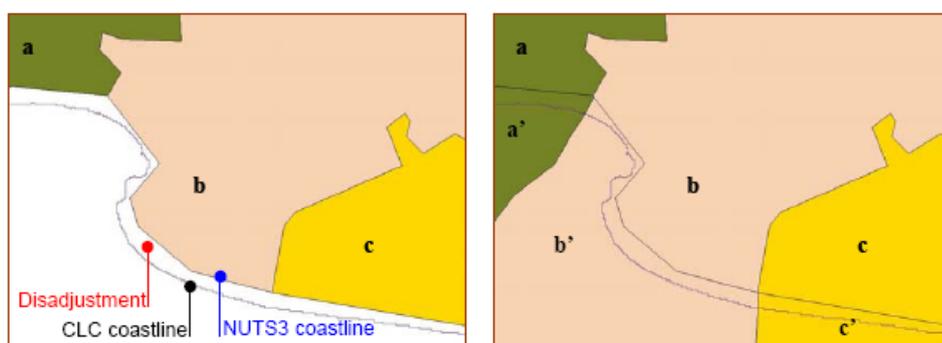


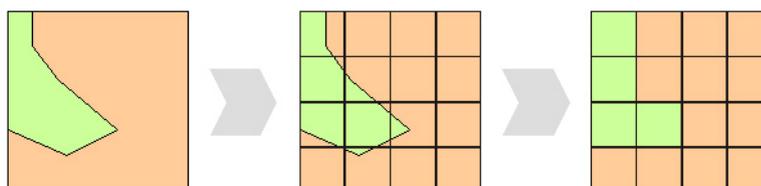
Figura 3. Proceso "Proximity"

De esta forma cada LARU se almacena en un campo de una tabla que posteriormente se vincula a la tabla LEAC_DATA a través del código de la celda de la malla de referencia. De esta forma, la vinculación entre ambas tablas nos permite consultar el modelo de datos y obtener los cambios o "stocks" en la cobertura del suelo a nivel de unidad territorial.

MÉTODOS DE INTEGRACIÓN DE INFORMACIÓN EN LA MALLA DE REFERENCIA

Dadas las posibilidades de integración de información utilizando la metodología LEAC, dependiendo de la naturaleza de cada variable o indicador que se quiera incluir en la base de datos se ha de definir el método más óptimo para llevarlo a cabo. Para ello, hemos definido y testeado con diferentes datos los siguientes métodos de integración:

Área Máxima: La celda de la malla de referencia hereda el valor de la unidad que cubre la mayor área de la celda. Este método es apropiado para variables incontables.



Cálculo proporcional: La celda hereda un valor calculado dependiendo de los valores de las unidades que caen en su interior, y su porcentaje de ocupación de la celda. Este método se plantea apropiado para variables contables.



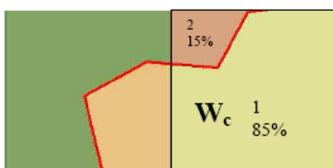
$$\text{Cell value} = \sum (V_i * \text{Share}_i)$$

Where: V_i = Value of unit i

Share_i = Share of unit i within the cell

In the example: $V_1 * 0.85 + V_2 * 0.15$

Cálculo proporcional y ponderado: La celda hereda un valor proporcional calculado, pero este valor está ponderado para cada celda acorde a una variable externa (por ejemplo, la población). Este método puede ser aplicado para hacer una mejor distribución territorial de indicadores socioeconómicos



$$\text{Cell value} = W_c \sum (V_i * \text{Share}_i)$$

Where: V_i = Value of unit i

Share_i = Share of unit i within the cell

W_c = weight assigned to cell c

In the example: $W_c * (V_1 * 0.85 + V_2 * 0.15)$

Dependiendo del tipo de indicador o variable que se quiera integrar en la malla de referencia se podrá definir un método más óptimo de integración. Independientemente del método de integración seleccionado, hay que considerar que ciertos indicadores vienen definidos por una unidad de área distinta al km². Este hecho hace que en ciertas ocasiones se haga necesaria una conversión de las variables considerando que el área de cada celda de la malla de referencia es de 1 km².

CLASIFICACIÓN DE LOS CAMBIOS: LAND COVER FLOWS

La metodología de LEAC considera los 1892 (43x44) posibles cambios entre las 44 clases existentes en *Corine Land Cover*. Los cambios en los usos del suelo están agrupados en lo que se denomina *Land Cover Flows* los cuales están a su vez clasificados de acuerdo a los siguientes procesos de usos del suelo:

- lcf1 Urban land management
- lcf2 Urban residential sprawl
- lcf3 Sprawl of economic sites and infrastructures
- lcf4 Agriculture internal conversions
- lcf5 Conversion from forested & natural land to agriculture
- lcf6 Withdrawal of farming
- lcf7 Forests creation and management
- lcf8 Water bodies creation and management
- lcf9 Changes of Land Cover due to natural and multiple causes

En la base de datos de LEAC los *Land Cover Flows* están descritos en la tabla FLATMATRIX. Esta tabla puede ser modificada de acuerdo a las posibles revisiones en la definición de los procesos de usos del suelo. Los cambios que se puedan producir en la definición de estos procesos, tiene un efecto directo en los resultados que se obtienen tras consultar los datos contenido en la base de datos LEAC.

EL MODELO DE DATOS DE LEAC

Una vez descritas los principales pasos a seguir de cara a asignar a la malla de referencia la información de la cobertura del suelo así como las unidades de análisis y de reporte, es importante ver cual es el modelo de datos final de LEAC y las posibilidades de consulta que brinda.

Es importante volver a resaltar que cada celda de la malla de referencia de 1km² esta asociada a los valores de cobertura del suelo así como a las unidades de reporte y análisis, LARUS y que dicha vinculación se establece a partir del código de la celda del GRID. El código de celda de la malla de referencia actúa como clave primaria y posibilita la vinculación de la tabla LEAC_DATA y la tabla de LARU. De esta forma el modelo de datos permite la consulta de estadísticas para un determinado LARU por medio de la selección de las celdas de la malla de referencia que se encuentran dentro de ellos. La base de datos de LEAC se compone de dos tablas principales y un conjunto de tablas de definición. Las tablas principales son las que se detallan a continuación:

LEAC_DATA: contiene la información de la cobertura del suelo para cada celda de la malla de referencia. Esta tabla a su vez puede vincularse con la tabla de LARUS y la tabla FLATMATRIX tablas que se utilizan para analizar los cambios en la cobertura del suelo.

LARU: Contiene los codigos para cada unidad de analisis y reporte para cada código de celda de la malla de referencia.

A continuación se muestra de forma gráfica cada una de estas tablas así como las relaciones que se establecen entre las tablas.

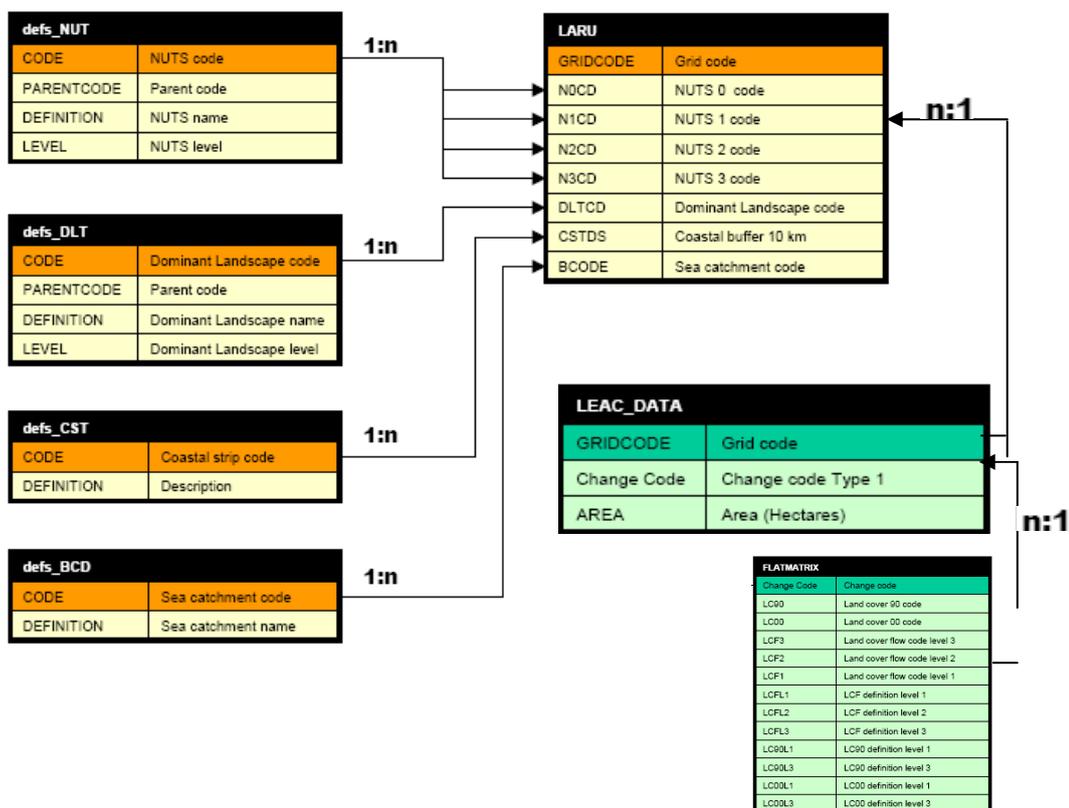


Figura 4. Modelo de datos de LEAC, tablas principales, relaciones y cardinalidad

CONCLUSIONES

Transformar la información espacial a una estructura clásica de modelo entidad-relación tiene el propósito de simplificar el procedimiento de geoprocésamiento de consulta de la información contenida en CLC. De esta forma, por ejemplo, consultas sobre la superficie de una determinada cobertura del suelo para una unidad de análisis resultado de la combinación entre diferentes unidades de análisis, se pueden general por medio de simple consultas SQL. Sin embargo, aunque por un lado la metodología LEAC facilita enormemente la consulta de estadísticas de usos del suelo, dado que la información espacial se convierte en información alfanumérica durante el proceso de construcción del modelo de datos, se pierde la capacidad directa que ofrece una herramienta SIG en explorar lo datos espaciales, haciéndose necesario la exportación de los resultados de las consultas a la base de datos LEAC y vincularlos a datos SIG para su posterior visualización y exploración visual.

En este sentido, el Centro Temático en Usos del suelo e Información Espacial ha comenzado un proyecto para el desarrollo de una aplicación Web que facilitará la consulta de la base de datos LEAC, y la posterior visualización de los resultados en forma de tablas, gráficos y mapas de coropletas que facilitará en gran medida la exploración de los resultados.

REFERENCIAS BIBLIOGRÁFICAS

Milego, R. and Ramos, M. (2010). Disaggregation of socioeconomic data into a regular grid: Results of the methodology testing phase, *ESPON Data Base*, Technical report.

Gomez, O., Paramo, F. (2005). Land and Ecosystems Accounts, Technical methodology, ETC/TE.

Weber, J. (2006). *Implementation of land and ecosystem accounts at the European Environment Agency*.

Land cover flows: <http://www.eea.europa.eu/data-and-maps/data/land-cover-flows-based-on-corine-land-cover-changes-database-1990-2000-1>