

Adaptación del Catálogo de Metadatos de la IDE Andalucía a la Directiva INSPIRE.

CASTILLA Eduardo, GAVIRA Manuel Andrés, PARDO Emilio, VILLAR Agustín

El Instituto de Estadística y Cartografía de Andalucía, como administrador de la Infraestructura de Datos Espaciales de Andalucía (IDE Andalucía), tiene encomendadas las tareas de mantenimiento, gestión y actualización del Geoportal y del Catálogo de Datos Espaciales. Ambas infraestructuras se ponen a disposición del público en el año 2006 y han ido actualizándose, tanto a nivel de software como de servicios prestados, a lo largo de estos años.

En lo que respecta al Catálogo de Datos Espaciales de Andalucía, se gestiona mediante el software OpenSource Geonetwork versión 2.4, contando en la actualidad con algo más de 45.000 registros, una cifra muy elevada, debido a que en origen se adoptó la decisión de generar los metadatos, de hojas de series cartográficas.

La aprobación de la Regla de Implementación de Metadatos de INSPIRE, con posterioridad a la publicación del catálogo de metadatos, ha provocado que estos no cumplan con los requerimientos planteados por ésta, por lo que se han tenido que abordar una serie de trabajos encaminados a modificar, tanto los metadatos en sí mismos, como la actualización del software base del catálogo.

Centrándonos en los metadatos, existían dos enfoques diferentes para abordar la actualización; "*Ex novo*" generando la totalidad de metadatos o actualizando los existentes a los requerimientos de la regla de implementación.

Esta segunda opción fue la elegida, debido al arduo trabajo realizado en la generación de los metadatos anteriores, con el inventariado de la información espacial generada en el ámbito de sus competencias de cada una de Consejerías de la Junta de Andalucía.

El trabajo de adecuación a la regla de implementación se ha abordado mediante el desarrollo varias aplicaciones que permiten la semi-automatización de los trabajos de transformación, actualización de contenidos y validación según el esquema de metadatos de INSPIRE.

La primera de las aplicaciones permite la transformación del esquema origen para ajustarlo al esquema de metadatos de INSPIRE y su posterior validación, mientras que la segunda de ellas permite el cambio, ya sea de forma individual o masiva del contenido de los metadatos.

Estos programas se han desarrollado con medios propios del Instituto de Estadística y Cartografía de Andalucía, en colaboración de los diferentes servicios que lo integran, si bien se ha tenido en cuenta su posible utilización por otros organismos que gestionen metadatos, integrados en la IDE Andalucía, por lo que se han definido opciones configurables que permite la adecuación de la transformación en función del esquema de metadatos de partida.

El presente artículo describe, tanto los trabajos previos de análisis de los metadatos existentes y la fase de desarrollo de las aplicaciones, como el funcionamiento de éstas.

PALABRAS CLAVE

Metadatos, Geonetwork, INSPIRE

ANTECEDENTES

La aprobación de la “Regla de Implementación de Metadatos de INSPIRE” y su aplicación en la IDE Andalucía ha traído como consecuencia la necesidad de adaptar el catálogo de Metadatos de Andalucía, en dos sentidos; por un lado el de cumplir con los objetivos que establece la regla, y por otro el de proceder a una revisión en profundidad el citado catálogo, por la desactualización que presenta y por el excesivo volumen de registros que contiene, debido a que en su origen fueron catalogadas todas las series cartográficas por hojas, llegando a principios del año 2014 a contar con algo más de 45.000 registros. Esta cantidad de datos trae como consecuencia los siguientes problemas de gestión del catálogo:

- Imposibilidad de actualizar metadatos de manera masiva, ya que el software utilizado en el catálogo de Andalucía (Geonetwork) no dispone de una herramienta eficaz para este fin.
- Alto consumo de recursos, tanto de personal, hardware como de tiempo, en cada proceso de actualización debido a la cantidad de datos con la que cuenta el catálogo.
- Inexistencia de software que sea capaz de generar metadatos conformes a la directiva INSPIRE. La aproximación más completa de las analizadas en el IECA es CatMDEdit¹, aún cuando los metadatos producidos con esta aplicación tampoco pasan el validador de INSPIRE de forma automática, siendo necesario el posterior trabajo de gabinete para su completa adaptación.

ANÁLISIS

Para la adaptación del catálogo a las necesidades antes descritas se procede al análisis del mismo, según una triple vertiente; análisis de la información existente, análisis del software Geonetwork y conclusiones sobre la información de partida y los objetivos propuestos.

- a) Análisis de los metadatos existentes. Los que hay actualmente en el catálogo de la IDE Andalucía están generados con el programa de creación de metadatos “MetaD”. Muchos de esos metadatos, al ser de hojas de series presentan un alto grado de similitud, cambiando únicamente la extensión geográfica y la fecha de creación del metadato.
- b) Para la Regla de implementación de INSPIRE, estos son datos redundantes y por tanto no cumplen con sus objetivos.
- c) Análisis del software Geonetwork. Geonetwork es un proyecto iniciado por la “Food and Agriculture Organization (FAO)” que proporciona servicios de gestión de información espacial. Está disponible su descarga en diferentes formatos para su posterior instalación de acuerdo a los requerimientos de cada usuario; una versión por defecto integrada con el servidor de aplicaciones Jetty y una versión únicamente con la aplicación, formato WAR, para integrar en aquellos entornos que ya disponen de servidores de aplicaciones. Igualmente está accesible su código fuente para su compilación que permite la adaptación por parte de los diferentes usuarios.

Las principales ventajas e inconvenientes de Geonetwork son:

1. Ventajas

- Edición en línea de metadatos con un sistema de plantillas.
- Funcionalidades de catálogo distribuido. Se puede trabajar con datos procedentes de diversos catálogos. La información de esos catálogos se puede sincronizar cada cierto tiempo contra la base de datos de Geonetwork, actualizándose así los cambios sobre el nodo central del sistema
- Compatible con la Geospatial Portal Reference Architecture, que es la recomendación del Open Geospatial Consortium para la creación y organización de geoportales
- Soporte para OGC-CSW 2.0.2 perfil ISO, OAI-PMH y protocolo Z39.50

- Dispone de herramientas para la gestión y control de acceso a la información
- Interfaz de usuario multilingüe.
- Adaptado a la directiva INSPIRE.
- Comunidad de desarrolladores activa e incluida en la Fundación OSGeo.
- Realiza búsquedas implementadas por el buscador LUCENE
- Implementado en Java EE
- Licencia GPL versión 2.

2. Inconvenientes.

- Se ha detectado que el programa Geonetwork, en su configuración por defecto presenta errores cuando se habilita la búsqueda según INSPIRE; no funciona correctamente ya que aun teniendo metadatos de series sólo localiza metadatos de conjuntos de datos y de servicios.
- Necesidad de compilar el código fuente si se quiere utilizar por defecto una vista alternativa.
- Manual desactualizado.
- No se pueden actualizar metadatos de manera masiva

- d) Conclusiones sobre la información de partida. Debido al gran volumen de información existente, el incumplimiento de los requerimientos de la directiva INSPIRE y el desajuste con las recomendaciones de la IDEE, hizo recomendable que la decisión a adoptar, desde la IDE Andalucía, fuera actualizar los metadatos, haciéndolos conformes a INSPIRE.

Una vez adaptado el catálogo a las nuevas necesidades, los nuevos metadatos se incluirán en un nuevo catálogo de Geonetwork (utilizando la versión 2.10.3) dado que el catálogo actualmente desplegado, no soporta los requerimientos de la directiva INSPIRE.

OBJETIVOS

A partir del análisis realizado, los objetivos generales del proyecto se centraron en dos aspectos fundamentales: Obtener un nuevo catálogo conforme a las especificaciones de INSPIRE, y desarrollar un método ágil de actualización y modificación de los mismos. Para la obtención de dichos objetivos, se plantearon tres bloques de tareas:

- I. Desarrollo de una aplicación para transformar los metadatos generados con MetaD, (aún cuando se deja abierta la posibilidad de tener soporte para otros programas), que validen con el esquema de INSPIRE y, por tanto, cumplan con dicha especificación. El desarrollado se ha llamado "Convalidación de Metadatos por INSPIRE" "CoMINSPIRE".
- II. Desarrollo de una aplicación de actualización de metadatos, ya sea de forma individual o masiva, que modifique la base de datos que administra Geonetwork, permitiendo la exportación a ficheros de diverso tipo (XML o PDF) o la importación a Geonetwork a partir de ficheros XML. Este software se ha llamado "Gestión de Metadatos de Andalucía", GesMAN.
- III. Adecuación de la interfaz de Geonetwork, para la publicación del catálogo en el Geoportal de la IDE Andalucía, que eviten desarrollos de interfaces intermedias, que complican el acceso a los datos.

Para el desarrollo de ambas aplicaciones se ha utilizado el lenguaje de programación Java y Eclipse como entorno de desarrollo. Para la adecuación de Geonetwork se han personalizado las interfaces alternativas que ofrece Geonetwork.

Flujo de trabajo desarrollado una vez desarrolladas las aplicaciones:

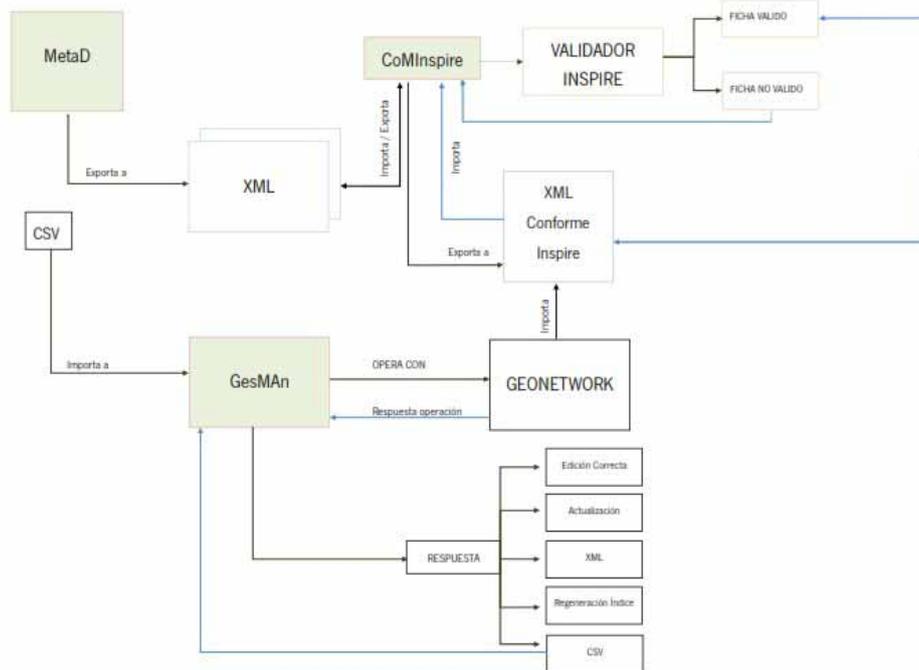


Figura 2: Flujo de trabajo en la Gestión de metadatos IDE Andalucía

I. Convalidación de Metadatos por INSPIRE. CoMINSPIRE

Objetivos: Con esta aplicación se pretende transformar los metadatos generados con “MetaD” y otras aplicaciones,

Figura 2: Pantalla de entrada de datos de la aplicación CoMINSPIRE

para que cumplan las especificaciones INSPIRE, de forma directa desde el programa, sin necesidad de visitar la Web de validación de INSPIRE. La interfaz de entrada del programa es la que se observa en la Figura 1

El programa presenta en la primera pantalla una serie de cajas de entrada de texto, con las siguientes funciones:

- Seleccionar la ruta de ubicación de los ficheros a transformar, ya sea un único o varios ficheros o un directorio completo.
- Seleccionar la ruta de ubicación de los metadatos transformados, en formato xml.
- Seleccionar el esquema xsl de transformación que va a utilizar el programa. Únicamente se habilitará si en el desplegable se opta por la opción "Personalizado". En este caso es necesario aportar la localización donde se encuentra el fichero del esquema (formato requerido xsl) que se utilizará en el proceso interno de transformación.
- Un selector con la opción de transformar metadatos de tipo "Conjunto de datos" o transformar Metadatos de "Servicios"
- Una caja de entrada de texto para escribir un identificador conocido. No es obligatorio.
- Dos desplegables en el que se escoge la palabra clave de INSPIRE (gmd:descriptiveKeywords) y el tema del metadato INSPIRE respectivamente (gmd:topicCategory).

Una vez introducidos los datos necesarios la aplicación realiza dos acciones:

1. Validación de metadatos. Esta acción permite validar datos directamente con la Web de validación de INSPIRE sin necesidad de abrir un navegador, con la posibilidad incluso de validar varios ficheros a la vez o un directorio entero.

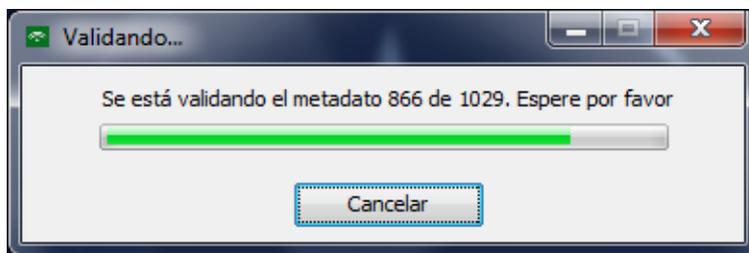


Figura 3: Pantalla de Información en el proceso de validación

La aplicación devuelve el mismo informe con el resultado de la validación que proporciona el validador del geportal INSPIRE.

2. Transformación de metadatos. Esta acción permite transformar aquellos metadatos que no hayan sido validados. Tras seleccionar los parámetros, se pulsa el botón "Transformar", iniciándose el proceso de transformación:

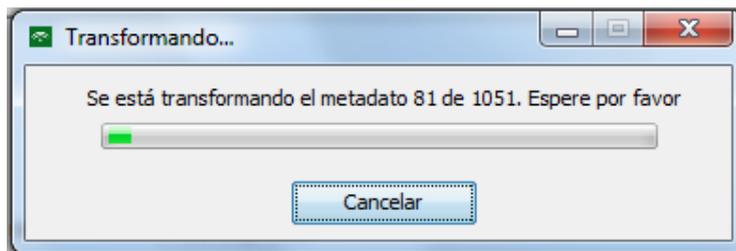


Figura 4: Pantalla de información sobre transformación de metadatos

II. Aplicación de actualización de metadatos (GesMAn)

El objetivo es crear un programa que sea capaz de:

- Editar, modificar y transformar metadatos uno a uno o de manera masiva.
- Exportar metadatos a XML o PDF en función de una o varias palabras clave, admitiendo cualquier valor susceptible de encontrarse en el metadato.
- Actualizar metadatos contenidos en ficheros XML.
- Creación y exportación de ficheros CSV para su posterior uso en la edición masiva de metadatos

Estas funcionalidades, o bien no las tiene Geonetwork, o si las tiene son difíciles de utilizar. En cualquier caso, siempre se trabajará con los datos de la tabla "Metadatos" de la base de datos de Geonetwork, en formato Postgres. Para la exportación a PDF se utilizará el servicio de generación que tiene incorporado Geonetwork.

El flujo de trabajo es el que se detalla a continuación. Al iniciar el programa se abre una ventana que solicita los datos de conexión a la base de datos:

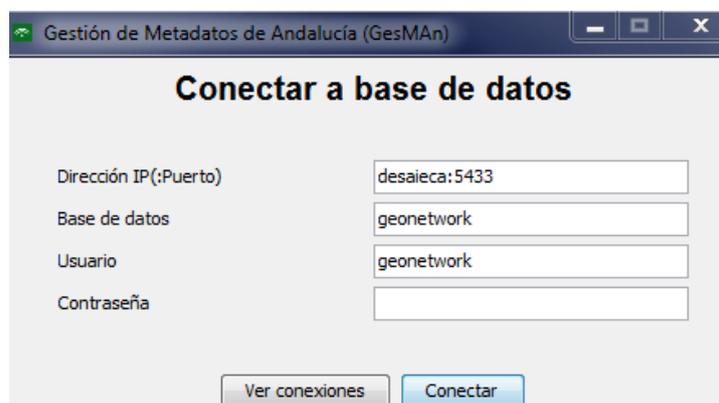


Figura 5: Pantalla de entrada de datos de la aplicación GesMan

Los elementos que contiene esta interfaz son:

- Cuatro cajas de entrada de texto, para escribir la dirección IP, el puerto, el nombre de la base de datos, el usuario y la contraseña.
- Un botón en el que se muestra una tabla con las conexiones que se hayan hecho anteriormente (si se pulsa por primera vez muestra un mensaje de error)
- Un segundo botón hace que ejecuta una conexión con los parámetros indicados.

Una vez que se haya realizado la conexión, aparece la ventana principal del programa, con 5 pestañas, cada una de las cuales tiene las siguientes funciones:

1ª Editar un metadato. Edición de metadatos individualmente. Para la selección del metadato de trabajo se identifican por los campos "Identificador" y "Título del metadato". Para que se pueda cambiar un registro tiene que existir un valor anterior y debiéndose escribir un valor posterior. Esta funcionalidad está pensada para la actualización de aquellos campos que pueden sufrir cambios constantes (Nombre Individual, organismo, teléfono....)

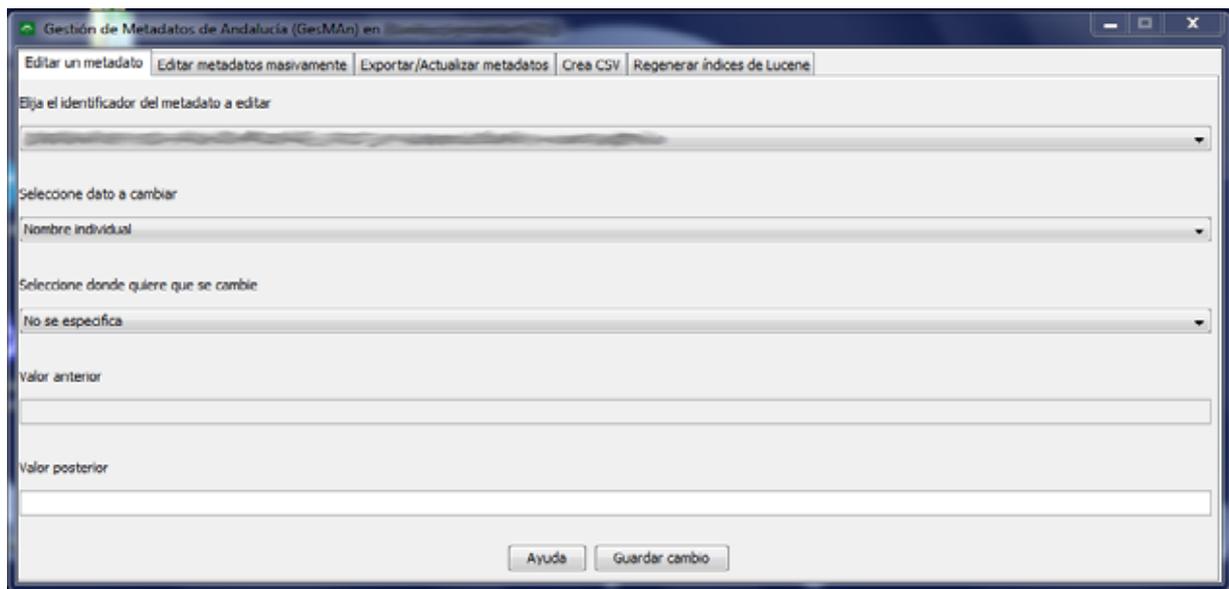


Figura 6: Pantalla de edición de un metadatos de la aplicación GesMAN.

2ª Editar metadatos masivamente. Su función es la edición masiva de metadatos, de aquellos campos que pueden sufrir cambios constantes (Nombre Individual, organismo, teléfono....). Proporciona una herramienta ágil para la gestión diaria de la Base de datos. Requiere disponer los datos en formato CSV.

Una vez seleccionado el fichero CSV, el programa requiere al usuario que le proporcione los identificadores de los metadatos a transformar, los datos a cambiar y los nuevos datos que los reemplazan.

Al pulsar el botón “Comenzar edición” se mostrará una pantalla que indica que se han cambiado los datos correctamente.

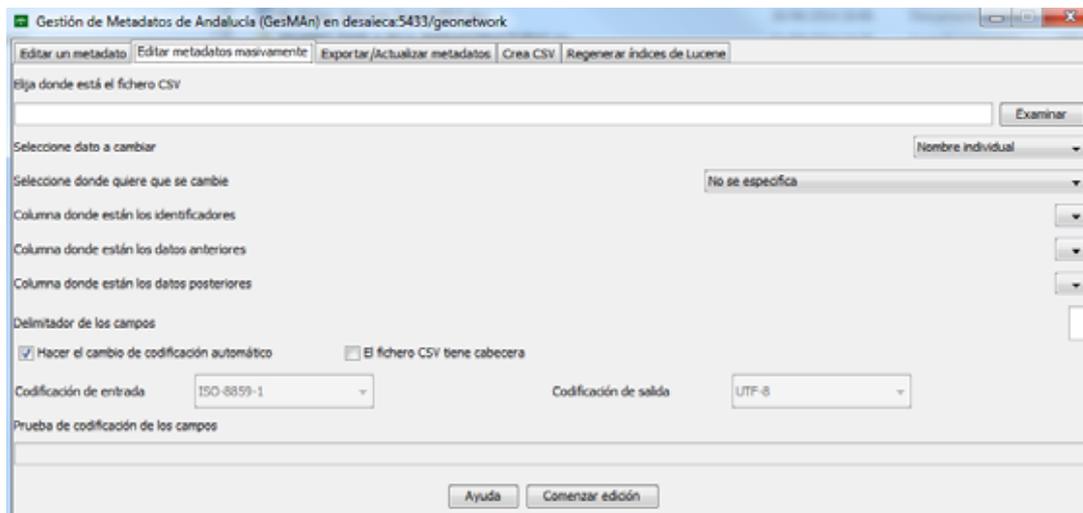


Figura 7: Pantalla de edición masiva de metadatos de la aplicación GesMAN

3ª Exportar/Actualizar metadatos. Esta pestaña nos permite exportar metadatos a XML o PDF y actualizar el metadato íntegro en la Base de datos de Geonetwork. Para realizar este proceso de cambio integral, es necesario que el metadato se encuentre en formato XML.

Para exportar metadatos, una caja de texto recibe las palabras (una o varias) que deben contener dichos metadatos. La interfaz permite seleccionar los metadatos del catálogo local o de otras accesibles en la Red. Igualmente se puede seleccionar el formato de exportación (XML, PDF).

Para actualizar los metadatos sólo se requiere indicar el lugar donde se encuentran los ficheros XML a actualizar. Cuando se pulsa el botón "Actualizar" la aplicación señala si se han actualizado correctamente, si ha habido algún error, o muestra una pantalla de advertencia en el caso que algunos metadatos no se hayan podido actualizar.

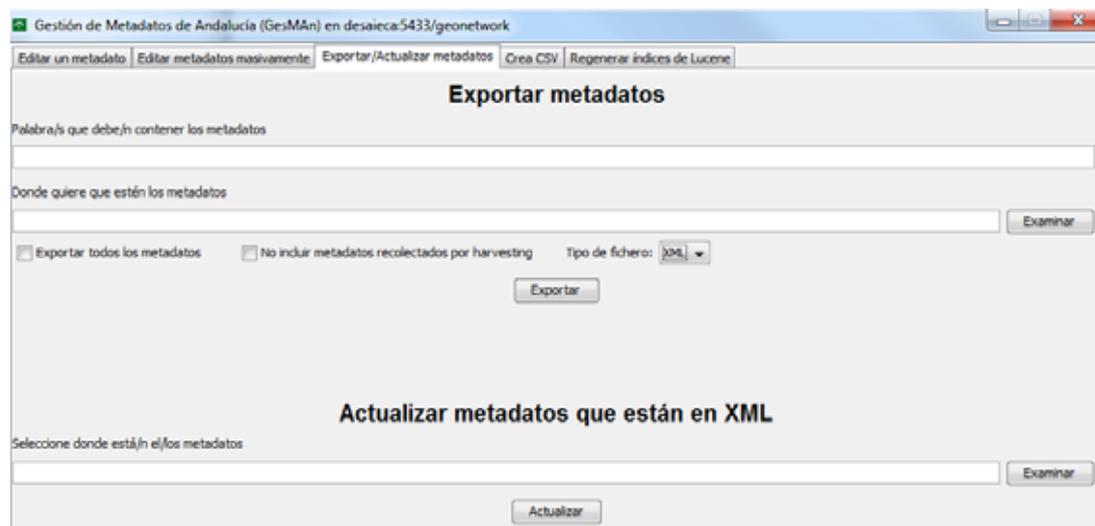


Figura 8: Pantalla de actualización completa de metadatos de la aplicación GesMAN

4ª Crea CSV. En esta pestaña se genera un fichero CSV con datos, que podrán ser utilizados posteriormente en la pestaña de edición masiva.

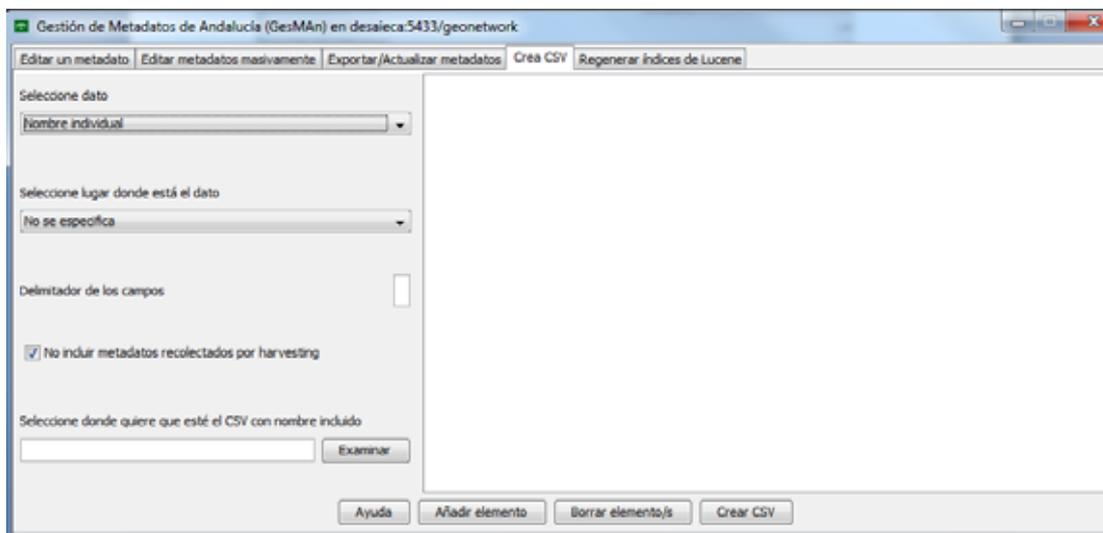


Figura 9: Pantalla de generación de ficheros CSV en la aplicación GesMAN

5ª Regenerar índices de Lucene. La última pestaña es para regenerar los índices de Lucene después de haber modificado los metadatos. En caso de no haber completado la totalidad de datos requeridos, saldría una pantalla de error. Esta funcionalidad facilitar la tarea de regeneración, sin necesidad de abrir un navegador.

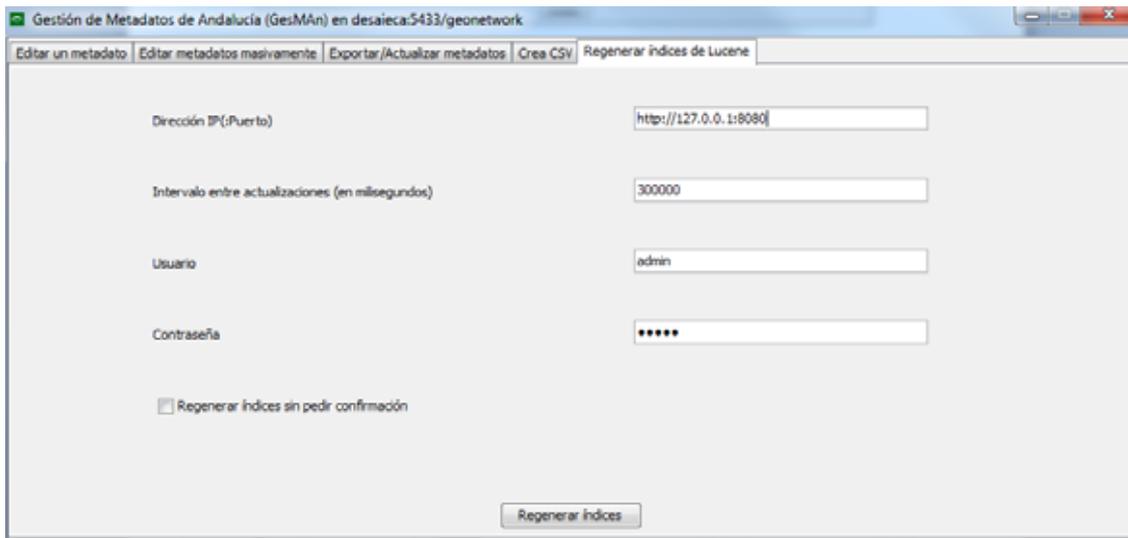


Figura 10: Pantalla de regeneración de índices de Lucene en la aplicación GesMAN

III. Modificación de Geonetwork

En la versión anterior del catálogo de metadatos, el módulo de consulta se desarrolló por medio de una interfaz web a medida, realizada en el lenguaje de programación PHP. Esto obligaba a hacer uso distintos lenguajes de programación en el servidor, Java para Geonetwork y PHP para la interfaz de consulta.

Para corregir esto, y aprovechando que las nuevas versiones de Geonetwork incluyen interfaces mas amigables que la versión por defecto (*search*, *searchtab*, *html5*), se ha desarrollado dentro del propio Geonetwork una interfaz simplificada que sustituye al cliente PHP, para la obtención de los resultados de las consultas.

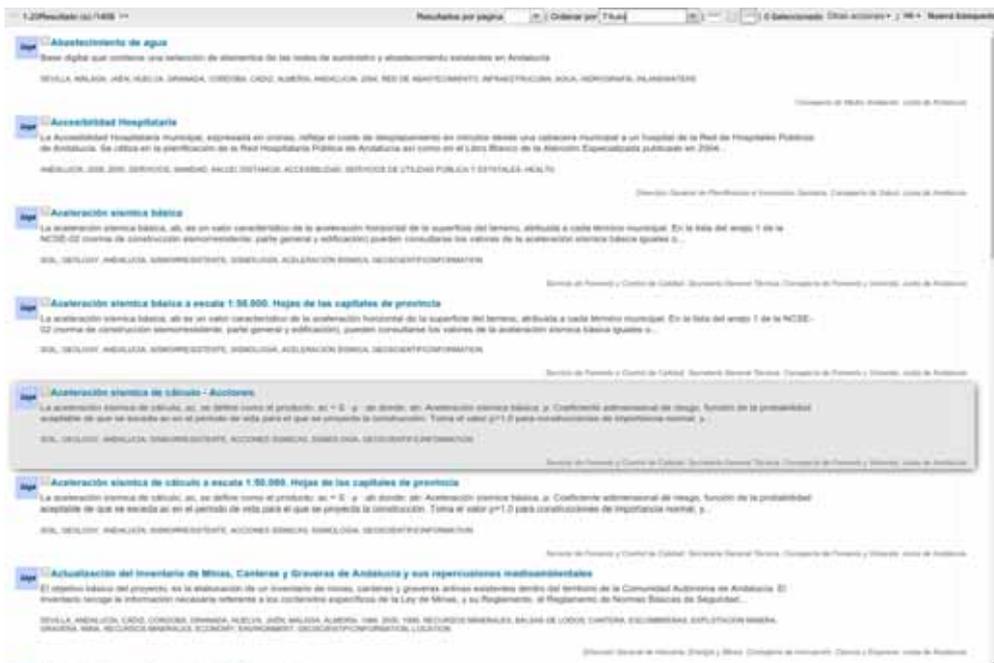


Figura 11 Módulo de búsqueda en la nueva configuración de Geonetwork.

Finalmente, se ha adaptado la visualización general de Geonetwork, manteniendo el uso de la interfaz por defecto para la gestión y actualización del catálogo.



Figura 12: Personalización de la visualización del programa Geonetwork.

CONCLUSIONES

La aprobación de la regla de implementación de metadatos de INSPIRE trajo como consecuencia la necesidad para el Instituto de Estadística y Cartografía de Andalucía de actualizar el catálogo de metadatos existente, tanto para adaptarse a los parámetros de la directiva INSPIRE, como en su actualización en número y calidad de los mismos. El procedimiento desarrollado por el IECA, en todos los casos realizado con medios propios, ha permitido la revisión y actualización de la totalidad del catálogo, así como la generación de un protocolo que permitirá en el futuro que los nuevos metadatos cumplan con la directiva europea.

El flujo de trabajo, de forma resumida es el siguiente:

- Generación de los metadatos con cualquier software específico y transformación y validación con el programa CoMINSPIRE.
- Publicación en el nuevo catálogo Geonetwork.
- Mantenimiento y actualización de los metadatos con el programa GesMAN.

Con estos pasos, el metadato ya es conforme a INSPIRE, se ha publicado en el catálogo y se muestra al público.

Además se ha conseguido que el entorno de actualización de metadatos, sea más sencillo y amigable de manera que puedan incorporarse al proceso de actualización de los mismos, personas que no tengan demasiados conocimientos sobre el entorno de trabajo de Geonetwork. Además el flujo de trabajo se ha hecho más flexible y los procesos de importación y exportación de datos de los catálogos se ha abierto a una pluralidad de formatos.

REFERENCIAS

[1] Geoportal Infraestructuras de Datos Espaciales de Andalucía. www.ideandalucia.es

[2] Catálogo de datos de la IDE Andalucía. www.ideandalucia.es/portal/web/ideandalucia/datos-y-servicios/catalogo-de-datos

[3] Catálogo de servicios de la IDE Andalucía. Catálogo de datos de la IDE Andalucía. <http://www.ideandalucia.es/portal/web/ideandalucia/datos-y-servicios/catalogo-de-servicios>

CASTILLA, Eduardo

email: eduardo.castilla@juntadeandalucia.es

Organización: Instituto de Estadística y Cartografía de Andalucía

Departamento: Servicio de Infraestructuras Geográficas.

PARDO, Emilio

email: emilioj.pardo@juntadeandalucia.es

Organización: Instituto de Estadística y Cartografía de Andalucía

Departamento: Servicio de Infraestructuras Geográficas.

GAVIRA, Manuel Andrés

email: mandres.gavira.ext@juntadeandalucia.es

Organización: Instituto de Estadística y Cartografía de Andalucía

Departamento: Servicio de Informática

VILLAR, Agustín

email: agustint.villar@juntadeandalucia.es

Organización: Instituto de Estadística y Cartografía de Andalucía

Departamento: Servicio de Infraestructuras Geográficas.

ⁱ Con posterioridad a la finalización del presente artículo, se ha publicado una nueva versión del programa, la 5.0. Ésta no ha sido analizada en lo concerniente a la validación con el esquema de la directiva Inspire.