



## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

### **ARMONIZACIÓN DE DATOS CONFORME AL MODELO INSPIRE. ANEXO I – LUGARES PROTEGIDO**

Mendive Lereñegui, Pedro <sup>(1)</sup>; Cardoso Santos, Juan Luis <sup>(1)</sup>; Cabello Martínez, María <sup>(1)</sup>; Echamendi Lorente, Pablo <sup>(1)</sup>; Alonso Pastor Del Coso, Fernando <sup>(2)</sup>.

<sup>(1)</sup> Tracasa; <sup>(2)</sup> Gobierno de Navarra.

A la hora de construir una Infraestructura de Datos Espaciales, la heterogeneidad de la información de partida donde cada proveedor de datos ya ha formulado enfoques específicos en su propio modelo de datos, y con frecuencia utilizando perfiles diferentes de metadatos, en el caso de que dispongan de estos últimos, se hace necesario desarrollar procesos de armonización con el fin de conseguir una interoperabilidad en la información disponible.

Esta ponencia recoge esta problemática a partir de un ejemplo real de transformación realizado en el marco de un proyecto europeo, en concreto con datos referidos a una temática incluida en el Anexo I de INSPIRE (Protected sites). Relataremos con detalle la experiencia de utilización de la herramienta Geoconverter, las dificultades encontradas y los resultados obtenidos en el proceso de armonización de datos.

El proyecto *NatureSDIplus*, en el que Tracasa ha participado como socio tecnológico, ha significado un primer paso en la construcción de una Infraestructura de Datos Espaciales para la Conservación de la Naturaleza, en el marco de la Unión Europea. El escenario de dicho proyecto nos pone ante la situación de varios proveedores de datos aportando información con diferentes modelos de datos, donde el proceso de armonización de dicha información se ha mostrado como imprescindible para demostrar la interoperabilidad de los conjuntos de datos y metadatos. Los trabajos de armonización se han llevado a cabo con éxito, desarrollando procedimientos para ello, analizando y aportando herramientas para realizar los distintos pasos a dar y complementándolo con metodologías de validación y controles de calidad. También cabe destacar el trabajo entrelazado con los grupos de trabajo de INSPIRE, testeando las diferentes versiones de los temas y/o aportando casos de uso a dichos grupos temáticos.

Este proyecto ha permitido obtener conclusiones extrapolables para futuras iniciativas dentro de este ámbito. La principal conclusión ha sido que los procedimientos implantados han conseguido que sea factible realizar procesos de armonización con datos provenientes de diversas fuentes. Pero no hay que obviar las dificultades tecnológicas y de capacitación que pueden existir entre los distintos proveedores de datos. También hay que señalar la dificultad de comprensión que puede existir de los modelos de INSPIRE y sobretodo la afección que puede producirse a la hora de remodelar los datos sobre los sistemas ya implantados en cada origen de datos.



## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

### ARMONIZACIÓN DE DATOS CONFORME AL MODELO INSPIRE. ANEXO I – LUGARES PROTEGIDOS

Mendive Lerendegui, Pedro <sup>(1)</sup>; Cardoso Santos, Juan Luis <sup>(1)</sup>; Cabello Martínez, María <sup>(1)</sup>; Echamendi Lorente, Pablo <sup>(1)</sup>; Alonso Pastor Del Coso, Fernando <sup>(2)</sup>.

<sup>(1)</sup>Tracasa. Dpto. de Sistemas de Información Territorial.

[pechamendi@tracasa.es](mailto:pechamendi@tracasa.es)

<sup>(2)</sup>Gobierno de Navarra. Dirección General de Medio Ambiente y Agua.

[fernando.alonso.pastordelcoso@cfnavarra.es](mailto:fernando.alonso.pastordelcoso@cfnavarra.es)

#### 1 La armonización de datos geográficos

El escenario actual en Europa por lo que respecta a los datos geográficos ha evolucionado enormemente en los últimos años. Es indudable que la Directiva INSPIRE, pero también otras cuestiones menos palpables quizás pero también relevantes, que tienen que ver con los nuevos comportamientos y ética de los usuarios, han supuesto un punto de inflexión y están ayudando a transformar profundamente la concepción que muchas organizaciones tenían de la información geográfica. El acceso a la información geográfica en estos momentos es mayor que nunca. No hace mucho aún, la viabilidad de un proyecto GIS dependía con frecuencia de la existencia y accesibilidad a unos datos, que no siempre se conseguía o que resultaban demasiado caros. Igualmente es de destacar el esfuerzo de documentación de metadatos que se está realizando en todas las organizaciones. Por lo tanto, si cada vez hay más datos, de más calidad y mejor documentados a disposición de los usuarios, se pone de relieve con mayor importancia que nunca la cuestión de la armonización de los datos. No resulta difícil hoy en día confeccionar un mapa de una zona fronteriza a partir de los datos disponibles a ambos lados de la frontera. Lo difícil es obtener una visión única y sintética que permita comparar la situación de uno u otro lado por las diferencias sintácticas y semánticas que existen en las leyendas.

INSPIRE ha decidido solucionar este problema, por lo menos a un cierto nivel, por medio de la armonización de datos. INSPIRE describe este proceso como la “acción de desarrollar un conjunto común de especificaciones de datos que posibiliten el acceso a datos espaciales a través de servicios de datos espaciales, de una manera que sea posible combinarlos con otros datos armonizados de una manera coherente. Nota: Esto incluye acuerdos acerca de sistemas de referencia espacial, sistemas de clasificación, esquemas de aplicación, etc.” [1]. Existen numerosos proyectos en la actualidad que están afrontando esta problemática en aras de lograr la máxima interoperabilidad de servicios y conjuntos y datos, entre ellos Nature-SDI *plus*, objeto de esta comunicación.



## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

### 2 El proyecto Nature-SDIplus

El establecimiento de la Red Natura 2000 [2] y los nuevos espacios fronterizos de la Unión Europea para la gestión de los espacios protegidos, han reforzado las sinergias entre la conservación de la naturaleza y la información geográfica. Ello ha generado la necesidad de una armonización entre todos los conjuntos de datos de manera que sean interoperables y accesibles a nivel de toda la Unión. Este mismo aspecto está íntimamente relacionado e incluido en la Directiva 2007/2/CE, más conocida como INSPIRE [3], cuyo objetivo es la creación de una Infraestructura de Datos Espaciales a nivel Europeo que sirva de apoyo para las políticas medioambientales en el continente.

El proyecto Nature-SDIplus [4] es una red de buenas prácticas para la creación de una infraestructura de datos espaciales sobre conservación de la naturaleza. Este proyecto fue seleccionado dentro de la convocatoria eContentplus de contenidos digitales para la Sociedad de la Información que realiza anualmente la Comisión Europea. En Nature-SDIplus han participado un total de 30 socios pertenecientes a 18 países de los 27 que forman la Unión Europea.

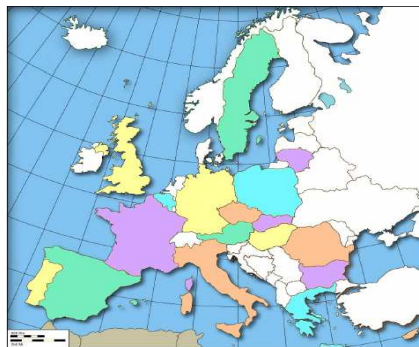


Figura 1. Países participantes en el proyecto Nature-SDIplus.

Los objetivos de la Red Nature-SDIplus se resumen en conseguir un conjunto de datos armonizados a nivel nacional que los haga más accesibles y explotables a través de metodologías innovadoras y ejemplos de buenas prácticas. En definitiva, involucrar a nuevos usuarios, compartir datos y buenas prácticas, mejorar y fomentar la explotación y reutilización de la información. Por consiguiente, podemos afirmar que este proyecto ha contribuido plenamente a la implementación de la Directiva INSPIRE, en concreto sobre la temática de Conservación de la Naturaleza, y más en concreto sobre los siguientes temas: Lugares protegidos (Anexo I), Regiones Biogeográficas, Hábitats y biotopos, y distribución de Especies (Anexo III).



El proyecto analizó la utilidad y accesibilidad de los datos disponibles por los proveedores de datos. Los socios proporcionaron una serie de datos representativos de los diferentes Estados Miembros de la Unión, y del análisis de los mismos, se pudo definir una propuesta multilingüe y multicultural común para un acceso más sencillo y estandarizado a dichos datos espaciales. El proyecto también ha aportado la creación de un nuevo Geoportal, basado en servicios Web OGC, que demuestra que es posible proporcionar accesibilidad a los datos a los diferentes grupos de usuarios y así involucrar a la ciudadanía en la conservación de la biodiversidad. Finalmente, también se tuvo en cuenta la difusión de los resultados del proyecto en diferentes actividades de formación y de diseminación.

Por resumir, las actividades del proyecto se agruparon en los siguientes paquetes de trabajo: WP1 *Project Management – Assessment and evaluation*; WP2 *User needs analysis and data policies*; WP3 *Interoperable datasets for nature conservation*; WP4 *Creation of the Nature-SDIplus portal*; WP5 *validation and testing of interoperable datasets*; WP6 *Nature-SDIplus Network development*; WP7 *Dissemination and awareness*.

En lo que respecta a nuestra participación en este proyecto, la Direc. Gral. de Medio Ambiente y Agua del Gobierno de Navarra se constituyó como un proveedor de contenidos, aportando varias capas temáticas relacionadas con la biodiversidad, y asesorando respecto a las funcionalidades del Geoportal. La participación de TRACASA se ha encuadrado en el marco de proveedores tecnológicos y expertos en implementación de Infraestructuras de Datos Espaciales, actuando como líderes del WP3: Interoperabilidad del conjunto de datos sobre Conservación de la Naturaleza y liderando igualmente las actividades TL 3.4 (procedimientos para la implementación del perfil de metadatos y del modelo de datos) y TL 4.1 (armonización de datos para el Geoportal Nature-SDIplus).

### **3 Consideraciones previas a un proceso de armonización: Análisis de la situación de partida**

Uno de los puntos de vista más importante a considerar en un proceso de armonización es el del proveedor de datos que necesita afrontar los trabajos. Desde Tracasa hemos trabajado de la mano con Gobierno de Navarra dando soporte tecnológico, solventando las dificultades encontradas a la hora de transformar los datos de partida para alcanzar el modelo de datos final.

En este apartado queremos mostrar algunos aspectos relevantes a tener en cuenta a la hora de afrontar un proceso de armonización, sobre todo el análisis de la información de partida pero también del modelo de datos que debemos llegar a cumplir. La experiencia obtenida en el WP2 del proyecto Nature-SDIplus sirve para resaltar estas cuestiones.



### 3.1 Análisis de los datos de partida.

Es necesario conocer las características de los datos de partida para conocer la cantidad de procesos previos que deberemos realizar a la información para que esta alcance la calidad mínima requerida. En el proyecto se analizaron los siguientes elementos:

Compatibilidad de los conjuntos de datos:

- Escala y resolución.
- Antigüedad de los datos.
- Formato.

Accesibilidad de los conjuntos de datos.

Cumplimiento de estándares y existencia de metadatos.

### 3.2 Análisis del modelo de salida: Especificaciones INSPIRE

En el tema del Anexo I - Lugares protegidos, objeto de esta comunicación, se decidió lógicamente utilizar el modelo de datos INSPIRE [5] como modelo de salida, debido a la coincidencia en el tiempo con la publicación de la versión definitiva de dichas especificaciones. Esto no fue posible, en cambio, para los temas del Anexo III, Biorregiones, hábitats y biotopos y distribución de especies. En estos casos, el trabajo desarrollado en el proyecto Nature-SDIplus sirvió para aportar experiencias y necesidades a los grupos temáticos de INSPIRE y los resultados se han visto reflejados en las últimas versiones que se han elaborado.

En el caso de la especificación de Lugares Protegidos, diremos que existían tres tipos de esquemas a los cuales acogerse: simple, completo y Natura2000. La disponibilidad de estos tipos, permitía el que los proveedores pudieran acogerse a unos u otros en función de la calidad de la información de partida. Los proveedores tuvieron que analizar en primer lugar los atributos obligatorios según estos esquemas para situarse y detectar posibles carencias en sus datos, que deberían corregir antes de realizar el proceso.

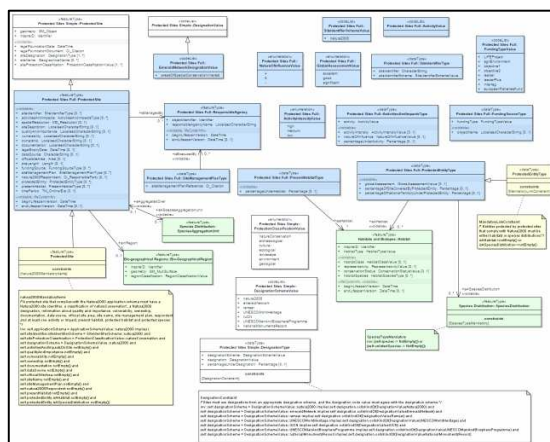


Figura 2. Modelo de datos INSPIRE para Lugares Protegidos

La heterogeneidad observada entre los conjuntos de datos de partida, y el análisis tanto de los modelos de origen como de salida obligan a una homogeneización para que en un sistema de explotación final todos los datos sean interoperables. Cumplir con este objetivo y a la vez definir unas pautas de cómo llevarlo a cabo es lo que se comúnmente se denomina un proceso de armonización de datos, que trataremos en el siguiente apartado.

#### 4 Proceso de armonización.

La armonización de los conjuntos de datos en un modelo común ofrece varias ventajas tanto a los proveedores de datos como a los usuarios, entre ellas, la posibilidad de acceder a la información de diferentes fuentes y combinar estos datos en una visión común. El objetivo es la interoperabilidad, definida en este proyecto como el proceso de “crear la posibilidad de combinar datos de orígenes heterogéneos en productos integrados, consistentes y no ambiguos, sin que el usuario final se de cuenta de ello”.

A la hora de decidir acerca de los métodos más adecuados para afrontar la armonización de datos espaciales, es importante distinguir entre los distintos tipos de heterogeneidad que pueden darse [6]: **Sintáctica**: Se refiere a las diferencias entre formatos. Es muy frecuente que se produzcan pérdidas de información al realizarse transformaciones entre formatos. **Estructural**: Se refiere a las diferencias entre esquemas, entre la aplicación de los modelos de datos conceptuales de cada proveedor. **Semántica**: Tiene que ver con el significado que un mismo término puede tener para dos proveedores de datos, que en función de su punto de vista o interés pueden ser muy diferentes. **Otros**: Diferencias en sistemas de coordenadas, idioma, etc.



Es preciso distinguir distintos niveles al hablar de armonización de datos espaciales [7], a nivel de modelo conceptual, de modelo lógico y de modelo físico.

El modelo conceptual es un modelo orientado a humanos de todo el mundo que consiste en objetos que un ser humano específico considere pertinente para un dominio específico. Se puede argumentar que sólo hay una realidad, mientras que hay modelos conceptuales casi tantos como personas hay. El modelo lógico se utiliza para explorar los conceptos de dominio y sus relaciones, y se expresa a menudo como modelos de clase en UML. Esta etapa es importante porque es el modelo lógico que se verá afectado por la agregación y la desagregación de los datos, ya que está armonizado. El modelo físico se utiliza para diseñar el esquema interno de una base de datos, que muestra cómo los datos físicos se almacenan en una máquina, es decir, las tablas de datos, las columnas de datos de las tablas y las relaciones entre las tablas almacenadas como archivos planos o bases de datos.

Como la armonización puede llegar a ser un tema complejo debido a que a veces no se tienen los conocimientos y las capacidades para analizar la información, para seleccionar las herramientas necesarias y seguir los pasos que se requieren, el proyecto Nature-SDI $plus$  decidió incluir la tarea de definir procedimientos y facilitar la elección de las herramientas a utilizar. El objetivo es que los proveedores de datos pudieran llevar a cabo con éxito el proceso de remodelación de los datos.

#### **4.1 Diseño de la propuesta metodológica**

Como se comentó previamente, el proyecto Nature-SDI $plus$  realizó una propuesta metodológica para realizar el proceso de armonización. Esta propuesta definió cinco fases, tal y como se ve en la imagen siguiente:

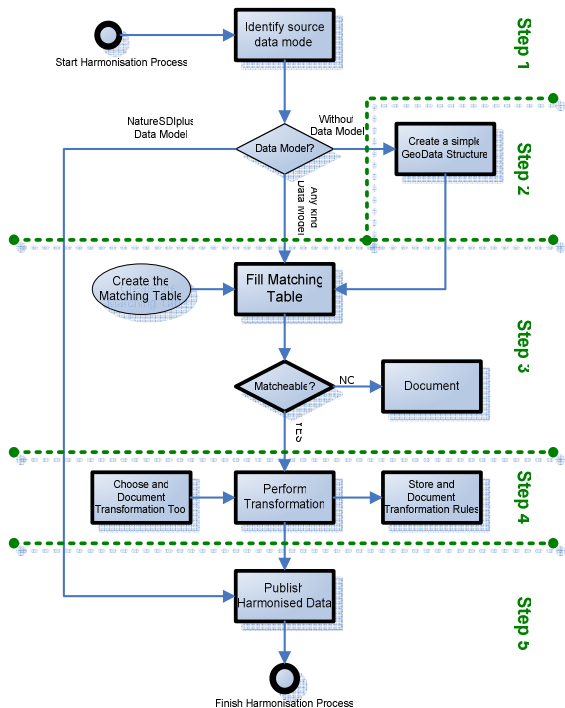


Figura 3. Diagrama del proceso de armonización

#### 4.1.1 Identificar y caracterizar la información (de origen y de salida).

Los proveedores de datos tienen que conocer los datos que poseen, y cómo se organizan y se almacenan. Si existe un modelo de datos, esto significa que los datos tienen alguna relación entre la información gráfica y alfanumérica, y si esta información está relacionada con los temas INSPIRE. Entonces, la información está lista para ser llenada en la tabla correspondiente. Cuanto más cerca del modelo fuente es el de destino, lógicamente, la transformación es más sencilla.

#### 4.1.2 Crear un mínimo modelo de datos (en caso de ser necesario)

Si no hay ningún tipo de modelo de datos, esto significa que los datos no tienen ninguna relación entre la información gráfica y alfanumérica. Será necesario crear una estructura de datos simple.





## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

### 4.1.3 Comparar la información final y de partida

La herramienta propuesta en el proyecto Nature-SDI*plus* para llevar a cabo esta tarea es la llamada *Matching Table* o matriz de correspondencias. Se trata de una tabla que permite comparar visualmente los datos y establecer relaciones entre los atributos de la información de partida y los atributos del modelo de datos final. Permite localizar la falta de información e incluso previsualizar cómo se va a generar parte de la información final. Esta matriz se ha considerado como una herramienta más dentro del proceso de armonización. En sí misma puede servir también para documentar problemas y necesidades encontradas por los proveedores de datos.

### 4.1.4 Elegir la herramienta de software para el remodelado.

En el proyecto Nature-SDI*plus* se estudiaron las capacidades y características de varias herramientas de software para llevar a cabo los procesos de remodelado de los datos. El resultado de este estudio facilitó que los proveedores de datos eligieran la herramienta que mejor se ajustaba a su situación particular. El siguiente paso fue realizar el trabajo de transformación de los datos propiamente dicho. Para los problemas encontrados se estableció un sistema de notificación por medio de unas plantillas (problemas con la información de partida, con el software, con los formatos, etc). Estas plantillas no sólo sirvieron para el seguimiento y resolución de problemas sino también para hacer una metodología de la calidad del proceso. En todo momento los resultados, problemas o dudas eran supervisados por un grupo técnico de armonización que se estableció como apoyo y garante del buen funcionamiento del proceso.

### 4.1.5 Crear los servicios de datos y catálogos para publicar la información.

Una vez que se obtuvieron los resultados y, tras un proceso de validación de los mismos, los datos armonizados se publicaron, bien directamente por los responsables del Geoportal o bien mediante Servicios Web creados por algunos de los socios del proyecto.

## 4.2 Selección de herramientas.

En el proceso de selección se valoraron diversas herramientas, algunas de ellas proporcionadas por los socios del proyecto y otras externas. En algunos casos se trataba de herramientas de código abierto y en otros de herramientas comerciales cuyos proveedores ofrecieron sin coste para apoyar los esfuerzos de armonización del proyecto Nature-SDI*plus*.

Para el análisis de las herramientas se creó una plantilla con preguntas sobre las características de cada herramienta. Estas características técnicas se pueden consultar en la figura siguiente:



## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

Tool Characteristics	Data Tools			
	HALE	Geomedia	GeoConverter	Snowflake GoPublisher
1. Enumerate input formats	*	*	*	*
2. Enumerate output formats	*	*	*	*
3. Read input data	Yes	Yes	Yes	Yes
4. Change Coordinate reference System	-	Yes	Yes	Yes
5. Allowing to define/load specific target model	Yes	Yes	Yes	Yes
6. Allowing manual /semiautomatic/ automatic mapping	Yes	Yes	Yes	Yes
7. Allowing to save transformation process	Yes	Yes	Yes	Yes
8. Any transformation quality control	No	Yes	No	No
9. Allowing multilingualism	No	Yes	No	No
10. License	LGPL	*	*	*
11. System Requirements	*	*	*	*

Figura 4. Características técnicas de las herramientas estudiadas.

El resumen final que se facilitó a los socios de proyecto Nature-SDIplus para realizar la armonización de sus datos detallaba los pasos a dar combinados con explicaciones técnicas de cada una de estas herramientas. De esta manera se conseguía facilitar no sólo el trabajo sino también la decisión de elegir una u otra solución tecnológica.

### 5 Ejemplo de armonización de datos de Lugares Protegidos con Geoconverter.

Tracasa, socio tecnológico del proyecto, aportó una aplicación de la suite GIS Geobide [8] denominada **Geoconverter**, que quedó a disposición de los proveedores de datos para que realizaran con ella las tareas de transformación de sus propios datos. La aplicación se utiliza para la conversión entre formatos de archivo de CAD/GIS. Para ello proporciona un asistente que nos guía para establecer los parámetros a conversión entre cualquiera de los formatos ofrecidos. Otras funcionalidades adicionales le permiten crear y editar atributos y establecer un sistema de coordenadas de referencia. Para poder cumplir con las necesidades de este proyecto se desarrolló una extensión de esta aplicación que incluía el soporte de los esquemas XSD que definen los modelos de datos INSPIRE así como el formato de salida GML 3.2.1.



El proceso de armonización de datos con Geoconverter realiza básicamente tres acciones: Transformar formatos (Geoconverter aporta la capacidad de lectura y escritura de más de 20 formatos distintos, entre ellos todos los que los proveedores de datos utilizaban); actualizar campos de la información según un modelo de datos consensuado y transformar sistemas de coordenadas de referencia.

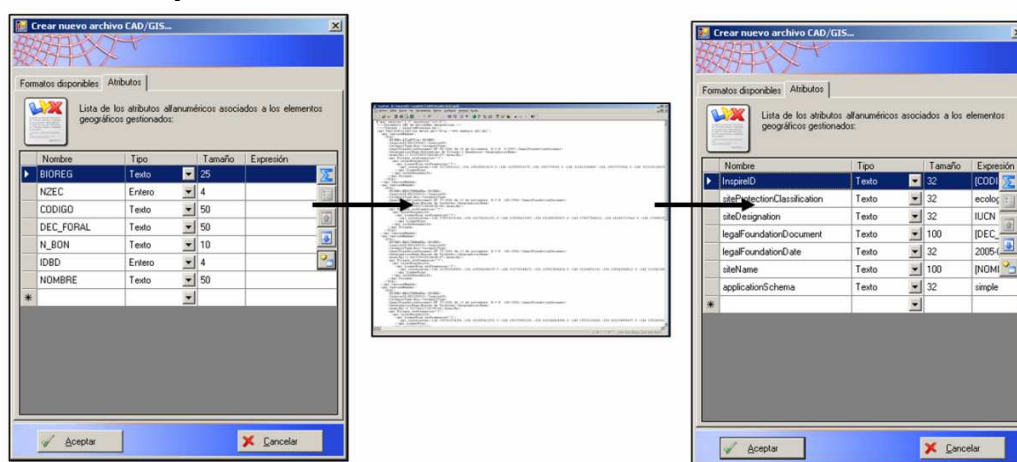


Figura 5: Interfaz de la aplicación Geoconverter.

El resultado obtenido de estos procesos fue una colección de ficheros conformes al modelo de datos Nature SDI plus para que los socios posteriormente pudieran publicarlos en forma de servicios OGC.

## 6 Conclusiones y lecciones aprendidas.

La experiencia adquirida nos ha mostrado que es posible conseguir el objetivo pero sin obviar las dificultades tecnológicas y organizativas que se pueden encontrar. Para ello, en proyectos como este, en el que hay tantos y tan heterogéneos participantes, se ha comprobado la importancia de establecer procedimientos de trabajo que sirvan de guía para que cada proveedor de datos sepa qué pasos dar en el proceso, siguiendo y analizando a la vez la calidad de los datos y procesos realizados. La valoración de los procesos y pautas establecidos nos ha hecho ver que la extrapolación de dichos procedimientos a otro tipo de procesos de armonización es posible, independientemente de su temática. En este caso otros proyectos europeos como Hlandata [9], que trabaja sobre la temática de coberturas y usos del territorio, ya se están apoyando en estas aportaciones.



## II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

No se pueden obviar las dificultades encontradas en el proceso de conseguir alcanzar el modelo INSPIRE, las cuales han permitido valorar el coste y necesidades a la hora de trabajar con los datos y también ha permitido testear e incluso aportar mejoras en los propios modelos desarrollados por la directiva INSPIRE. Es vital insistir en la necesidad de formación en metodologías de armonización y en la utilización de las herramientas. Además, los proveedores de datos también deberán contemplar la reingeniería que en algunos casos se requerirá para llegar al objetivo final de armonización, haciendo que modelos de datos y sistemas ya en funcionamiento, a nivel local, regional o nacional, deban ser modificados o duplicados, con lo que ello puede suponer.

Otra conclusión es la posibilidad que nos ha brindado el participar en un proyecto de este tipo para estudiar y conocer características técnicas de herramientas de software y en nuestro caso en concreto para testear, mejorar y divulgar la utilización de la herramienta Geoconverter para la realización de procesos de reales de armonización.

La armonización puede ser una operación más o menos compleja, dependiendo del modelo de datos y de la especialización del equipo de trabajo. La experiencia obtenida de la utilización de Geoconverter en este proyecto demostró que con una pequeña tutorización, es muy sencillo realizar este tipo de tareas, incluso en pequeñas organizaciones sin personal GIS altamente cualificado.

### 7 Referencias bibliográficas

- [1] D2.3: Definition of Annex Themes and Scope, Version 3.0. [http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3\\_Definition\\_of\\_Annex\\_Themes\\_and\\_scope\\_v3.0.pdf](http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pdf)
- [2] Red natura2000. <http://www.rednatura2000.info/>
- [3] Directiva europea INSPIRE. <http://inspire.jrc.ec.europa.eu/>
- [4] Proyecto Nature-SDIplus. <http://www.nature-sdi.eu/>
- [5] [http://inspire.jrc.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_PS\\_v3.1.pdf](http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.1.pdf)
- [6] Breu, A.: Data Harmonisation Topic. Data Remodelling. GIS4EU Newsletter n. 6. (2009).
- [7] Mendive, P., Cardoso, J. L., Cabello, M.: Nature-SDIplus Deliverable 3.5 "Procedures for metadata profile and data model implementation" (2010).
- [8] Suite GIS Geobide. <http://www.geobide.es>
- [9] Proyecto Hlandata. <http://www.hlandata.eu>