

Harmonização de Dados Geográficos INSPIRE usando o HALE: Exemplo de aplicação a dados geográficos do tema Sítios Protegidos

FURTADO, Danilo¹; GOMES, Ana Luisa²; FONSECA, Alexandra³;

dfurtado@igeo.pt; luisa.gomes@igeo.pt; afonseca@igeo.pt;
Instituto Geográfico Português

Resumo

As Infra-estruturas de Dados Espaciais (IDE) possibilitam o acesso a grandes quantidades de informação espacial e a ferramentas de visualização e exploração dessa informação. A referida informação provém de diferentes organizações, com diferentes formas de representar a informação geográfica, diferentes formatos e Sistemas de Referência Espacial, tornando a integração de dados uma tarefa complexa. A Directiva INSPIRE pretende conduzir à harmonização de dados provenientes de diferentes fontes, definindo-a como “o processo que permite desenvolver especificações para conjuntos de dados geográficos, de modo a que seja possível aceder a estes dados através de serviços, numa representação que permita combinar esses dados com outros dados, harmonizados de forma coerente”. Têm vindo a ser desenvolvidos procedimentos e ferramentas de harmonização de dados geográficos com o intuito de facilitar o seu intercâmbio e exploração integrada. Através de interfaces gráficas adequadas essas ferramentas permitem a realização pelo utilizador de diferentes tarefas, como por exemplo, o mapeamento de esquemas conceptuais (conceptual *schema*), a conversão de sistemas de referência, a concretização de operações de *edge matching*. Este artigo descreve o processo de harmonização de dados geográficos relativos ao tema Sítios Protegidos do Anexo I da Directiva INSPIRE, utilizando a ferramenta HALE - HUMBOLDT *Alignment Editor*, desenvolvida no âmbito de um projecto financiado pela Comissão Europeia.

Palavras-chave: INSPIRE, IDE, Harmonização de dados geográficos, Sítios Protegidos

1. Introdução

O recurso à informação geográfica tem vindo a desempenhar um papel fundamental no desenvolvimento de actividades no domínio do ambiente e da conservação da natureza. No entanto, existem grandes problemas no acesso e utilização deste tipo de informação, tanto a nível europeu como nacional, nomeadamente, duplicação de informação, incompatibilidade de formatos entre conjuntos e serviços de dados geográficos e entraves à sua partilha e reutilização, resultando na falta de interoperabilidade técnica e semântica.

As Infra-estruturas de Dados Espaciais (IDE) podem contribuir para colmatar as necessidades de informação e contribuir para uma melhor utilização de conjuntos de dados geográficos, ao possibilitarem o acesso a grandes quantidades de informação espacial e a ferramentas de visualização e exploração dessa informação. A referida informação provém de diferentes organizações, com diferentes formas de representar a informação geográfica, diferentes formatos e Sistemas de Referência Espacial (SRS), tornando a integração de dados uma tarefa complexa.

A Directiva INSPIRE (Directiva 2007/2/EC do Parlamento Europeu e do Conselho de 14 de Março de 2007) tem como objectivo principal conduzir à criação da Infra-estrutura Europeia de Dados Espaciais para apoio à definição e aplicação das políticas ambientais na Europa Comunitária. A directiva incide sobre informação espacial referente a um conjunto de temas distribuídos por três anexos que abrangem dados espaciais de natureza trans-sectorial e dados espaciais específicos do sector ambiental [1].

Ao obrigar os Estados Membros a adoptarem um conjunto de regras ou disposições de execução relativas às diferentes componentes das respectivas infra-estruturas de informação geográfica (e.g. metadados, conjuntos de dados geográficos, serviços de dados geográficos, serviços de rede) a directiva pretende contribuir para ultrapassar os problemas de integração de dados geográficos anteriormente referidos.

A harmonização de conjuntos de dados geográficos (CGD) é um dos aspectos mais críticos da implementação da directiva, requerendo um esforço significativo por parte das instituições produtoras de dados geográficos dos diferentes Estados Membros para alterarem os seus procedimentos no sentido de produzirem informação harmonizada de acordo com as especificações de dados INSPIRE. Estas instituições necessitam de apoio em termos técnicos e organizacionais para cumprirem este objectivo.

Com o intuito de suportar os desenvolvimentos necessários à aplicação da Directiva têm surgido diversos projectos Europeus centrados nas temáticas INSPIRE, alguns de natureza transversal, estudando por exemplo o desenvolvimento de arquitecturas para a harmonização de dados e integração de

serviços ou metodologias e ferramentas para suporte aos processos de especificação de dados e outros de natureza específica, como por exemplo o estudo da implementação das especificações de dados associadas a determinados temas dos Anexos [2].

Foi desenvolvida no âmbito de um desses projectos, o projecto HUMBOLDT-Towards the Harmonisation of Spatial Information in Europe [3], uma ferramenta de harmonização de dados, HALE - HUMBOLDT Alignment Editor, que permite de forma gráfica definir o mapeamento entre conceitos nos esquemas conceptuais (conceptual schema) dos CDG. A sua aplicação a dados geográficos do Tema “Sítios Protegidos” foi concretizada no âmbito de um outro projecto designado por NatureSDIplus – Best Network for SDI in Nature Conservation.

Este artigo descreve o processo de harmonização de dados geográficos relativos ao Tema Sítios Protegidos do Anexo I da Directiva INSPIRE, utilizando esta ferramenta.

2. Directiva INSPIRE: harmonização de dados geográficos

A Directiva INSPIRE (2007/2/EC) pretende viabilizar a disponibilização de conjuntos e serviços de dados geográficos de acordo com princípios e regras comuns, com o intuito de permitir a qualquer utilizador pesquisar, visualizar e aceder de um modo interoperável, a informação geográfica proveniente de diferentes Estados Membros.

A directiva incide sobre informação espacial referente a um conjunto de temas distribuídos por três anexos que abrangem dados espaciais de natureza trans-sectorial e dados espaciais específicos do sector ambiental. No que concerne directamente ao domínio da Conservação da Natureza a directiva considera temas diversos como o tema Sítios Protegidos pertencente ao Anexo I ou os temas Habitats e Biótopos, Regiões Biogeográficas e Distribuição de Espécies incluídos no Anexo III.

A entrada em vigor da Directiva INSPIRE implica o cumprimento pelos Estados Membros de um conjunto de disposições de execução que constituem regras de implementação a aplicar no contexto das diferentes componentes das respectivas infra-estruturas (e.g. metadados, CDG, serviços de dados geográficos, serviços de rede).

De acordo com a Directiva INSPIRE a harmonização de dados é entendida como “o processo que permite desenvolver especificações para conjuntos de dados geográficos, de modo a que seja possível aceder a estes dados através de serviços, numa representação que permite combinar esses dados com outros dados harmonizados de forma coerente”.

Os dados geográficos apresentam problemas de compatibilidade a diversos níveis nomeadamente: diferentes sistemas de coordenadas; desalinhamento de limites gerando lacunas e sobreposições; diferentes modelos de dados; diferentes perfis de metadados.

A harmonização de dados geográficos implica que sejam aplicadas regras aos esquemas conceptuais (“conceptual schema”) dos modelos de dados dos conjuntos de dados geográficos com o objectivo de proceder à sua transformação no modelo de dados INSPIRE. O esquema conceptual é uma descrição rigorosa do modelo conceptual que descreve o universo associado aos dados.

As disposições de execução que definem as especificações para os dados relativos aos temas do Anexo I [4] estão disponíveis desde Dezembro de 2010 devendo ser aplicadas para os novos dados até dois anos após essa data. Estas especificações definem um modelo de dados comum para os temas do Anexo I da Directiva: Sistemas de referência, Sistemas de quadrículas geográficas, Toponímia, Unidades administrativas, Endereços, Prédios, Redes de transporte, Hidrografia e Sítios protegidos.

3. Especificações de dados INSPIRE para o Tema “Sítios Protegidos”

As especificações de dados INSPIRE para o tema do Anexo I Sítios Protegidos estão descritas no documento INSPIRE Data Specification on Protected Sites. Guidelines v 3.1.0 (03.05.2010) [4], que identifica três perfis de aplicação, cada um com finalidades diferentes:

- Perfil Simples (*Simple Application Schema*) - representa o CORE, com um número limitado de atributos, tais como: geometria, identificador, nome, tipo de designação, fundamentos legais, data e documentos de referência;
- Perfil Completo (*Full Application Schema*) – inclui todos os atributos considerados relevantes para o tema “Sítios Protegidos”, nomeadamente atributos relacionados com factores temporais, sendo a maior parte opcional;
- Perfil Natura 2000 (*Natura 2000 Schema*) - inclui todos os atributos do perfil completo mas tendo como atributos obrigatórios os exigidos aos Estados-Membros para a actualização e manutenção de dados da Rede Natura2000.

A harmonização de dados de acordo com as especificações de dados INSPIRE implica o estabelecimento de correspondências de atributos entre os *schemas* dos CDG e as especificações INSPIRE. A Figura 1 ilustra esse processo para um CDG nacional – Limites das Áreas Protegidas.

Anexo I
MACTHING TABLE: APS_2009

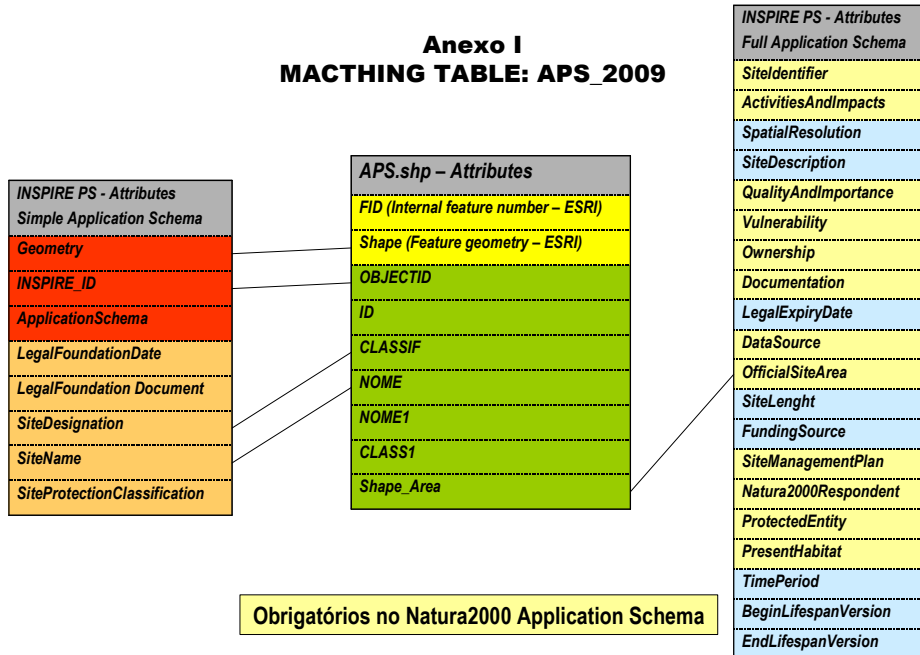


Figura 1. Estabelecimento de Correspondências entre os atributos do CDG – Limites das Áreas Protegidas e as especificações de dados INSPIRE do tema Sítios Protegidos

4. Harmonização de Dados Geográficos usando o HALE

A entrada em vigor da Directiva INSPIRE, conduziu ao surgimento de numerosos projectos Europeus centrados nas suas temáticas, como o são os projectos HUMBOLDT, GIS4EU, EURADIN, NATURE-SDIplus ou BRISEIDE, uns de natureza transversal, outros focados em temáticas específicas como é o caso da conservação da natureza.

É no contexto do projecto HUMBOLDT que surge a ferramenta de harmonização de dados geográficos cuja aplicação é discutida neste artigo.

O projecto HUMBOLDT (<http://www.esdi-humboldt.eu/home.html>) centrou-se no desenvolvimento de uma arquitectura para a harmonização de dados e integração de serviços nos termos da directiva INSPIRE considerando os objectivos do programa Global Monitoring for Environment and Security (GMES). Os trabalhos no âmbito do projecto foram desenvolvidos em estreita ligação com os grupos de trabalho responsáveis pelas especificações INSPIRE.

Pretendendo contribuir para a harmonização de informação espacial na Europa no contexto de uma IDE Europeia, o referido projecto visou o desenvolvimento de

metodologias e ferramentas que conduzissem à redução dos custos associados ao esforço de implementação da directiva INSPIRE.

Assim, as actividades do projecto pretenderam: (1) suportar a integração de serviços e dados para facilitar a tomada de decisões em aplicações nas áreas da segurança e do ambiente; (2) dar suporte a programadores e utilizadores no desenvolvimento de programas que interajam com a IDE Europeia; (3) Constituir uma plataforma para teste das regras de implementação INSPIRE; (4) criar especificações técnicas que melhorem as regras estabelecidas no âmbito do INSPIRE; (5) desenvolver e implementar novas metodologias e ferramentas para suporte aos processos de especificação de dados e de harmonização.

Os cenários constituíram uma componente essencial ao desenvolvimento do HUMBOLDT Framework uma vez que o projecto visou responder a necessidades de áreas de aplicação diversas. O projecto incluiu assim a definição de oito cenários que aplicam as ferramentas e metodologias criadas testando-as em condições realistas e em diversas áreas de aplicação (e.g. gestão florestal, gestão de áreas protegidas gestão de risco de cheias).

Os cenários usam dados de temas incluídos nos anexos da directiva INSPIRE e aplicam as respectivas regras de implementação. Para cada cenário é criado um modelo de dados comum tendo em conta o INSPIRE Generic Conceptual Model[5] e a Methodology for the development of data specifications: baseline version [6] entre outros. Os cenários foram usados para testar a exequibilidade dos processos de harmonização de dados.

O cenário das Áreas Protegidas (HS Protected Areas Scenario) pretende analisar a aplicabilidade do HUMBOLDT Framework à temática da Conservação da Natureza. O envolvimento do Instituto Geográfico Português (IGP) no projecto incidiu de forma significativa no contributo para este cenário.

Neste contexto e com a colaboração do Instituto de Conservação da Natureza e Biodiversidade (ICNB), foram identificados os Conjuntos de Dados Geográficos nacionais que se enquadram no tema Sítios Protegidos e por associação nos outros temas associados à Conservação da Natureza, todos incluídos no Anexo III.

Tabela 1 – CDG do domínio da Conservação da Natureza harmonizados com o HALE.

CDG	ANEXO da Directiva
APS – Limite das Áreas Protegidas	I – Sítios Protegidos
ZPE – Zona de Protecção Especial (Directiva Aves)	I – Sítios Protegidos
Sítios Classificados (Directiva Habitats)	I – Sítios Protegidos
HABITAT 1110	III - Distribuição de Habitats

Triturus_marmoratus (Anfíbios)	III – Distribuição de Espécies
Alosa_alosa (Peixes)	III – Distribuição de Espécies
Coluber_hippocrepis (Répteis)	III – Distribuição de Espécies
Lutra-lutra (Mamíferos)	III – Distribuição de Espécies
Merops_apiaster (Aves)	III – Distribuição de Espécies
Raculeatus (Plantas)	III – Distribuição de Espécies

Estes dados foram harmonizados no contexto do projecto NatureSDIplus – Best Network for SDI in Nature Conservation [7], utilizando a ferramenta HUMBOLDT Alignment Editor – HALE desenvolvida no âmbito do projecto HUMBOLDT[3], uma aplicação que permite de forma gráfica a definição de correspondências entre *conceptual schemas*, bem como a definição de transformações entre atributos desses *schemas*. Trata-se de uma ferramenta de suporte à harmonização de dados geográficos que permite de forma gráfica definir o mapeamento (correspondências) entre conceitos nos *conceptual schema* dos CDG e definir transformações entre atributos desses *schema*.

Esta aplicação faz parte de um conjunto de aplicações desenvolvidas no âmbito do HUMBOLDT Framework (Figura 2) com o objectivo de tornar os CDG dos Estados Membros harmonizados e interoperáveis de acordo com as especificações dos modelos de dados da Directiva INSPIRE.

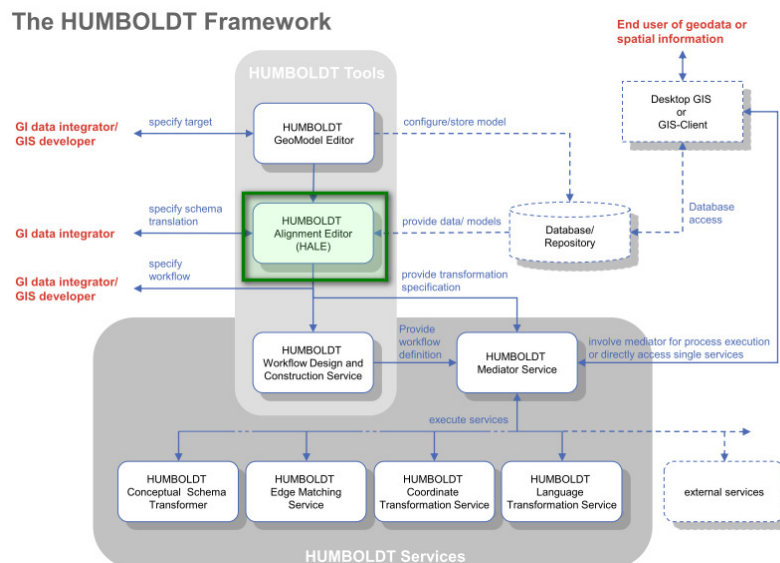


Figura 2. Arquitetura do HUMBOLDT Framework.

O HALE é uma aplicação Eclipse RCP, multiplataforma que funciona nas versões do Linux, GTK, bem como Mac OsX 10.5+ e MS Windows. Encontra-se actualmente na versão 2.1.1 de 27 /06/2011. Possui uma licença de distribuição de código aberto (Open Source) sobre a versão GNU Lesser General Public License 3 (LGPL v3).

A aplicação HALE permite resolver diversos problemas relacionados com a interoperabilidade:

- Diferenças entre *conceptual schema* - a aplicação permite definir regras de mapeamento para as classes e atributos entre o source *conceptual schema* e o target *conceptual schema*;
- Diferenças nos metadados - permite criar regras de mapeamento para os elementos dos metadados;
- Inconsistências nos aspectos espaciais e temporais dos dados - possui funcionalidades que permitem a transformação geométrica dos dados;
- Representações múltiplas - permite definir regras para lidar com múltiplas representações de um objecto em termos de precedências.

A descrição do fluxo de trabalho inclui um conjunto de passos necessários à preparação do esquema de transformação usando HALE.

1. Carregamento dos *schemas* e dos dados: *source e target schemas*

O primeiro passo é carregar o *source e target schemas* no HALE Schema Explorer. Esta versão permite o carregamento de qualquer XML schema, GML (versões 2.1, 3.1 e 3.2) bem como ESRI ShapeFiles (*.shp). Os schemas podem também ser carregados a partir de Web Feature Service (WFS) através da operação GetCapabilities.

Neste caso de estudo foi utilizado o ESRI ShapeFile das Áreas Protegidas como *source schema* e o INSPIRE *Protected Sites Full schema* como *target schema*.

Após carregamento dos *schemas* é possível carregar os dados de origem para a aplicação. Esta operação possibilita a visualização da representação cartográfica dos dados na Map Viewer. É possível dividir a Map Viewer e visualizar os dados de origem e a respectiva transformação lado a lado. Estão disponíveis funções de navegação e criação de simbologia. Se o sistema não consegue identificar claramente o CRS (Coordinate Reference System) a partir dos dados, é solicitado ao utilizador o fornecimento através do código EPSG ou do Well-Known Text para o Sistema de Referência utilizado.

2. Mapeamento dos *schemas*

O passo seguinte corresponde ao mapeamento entre as entidades do *source* e *target schemas* usando a tabela de correspondências mencionada anteriormente. A Figura seguinte ilustra o mapeamento entre os *source* e *target schemas*.

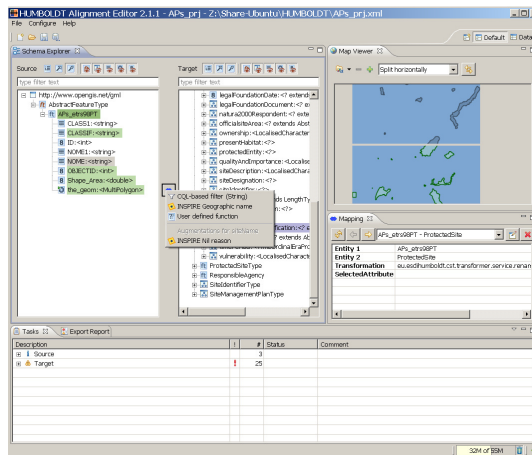


Figura 3. Mapeamento entre o Source e o Target schemas no HALE schema explorer

Neste caso de estudo foram utilizadas as seguintes funcionalidades do HALE no processo de harmonização:

- Retype – função que permite estabelecer a relação geométrica entre os *source* e *target schemas*;
- Attribute Rename – permite renomear os campos dos atributos alfanuméricos;
- INSPIRE Identifier – cria o atributo IdentifierPropertyType, que permite passar a ter os ID dos dados em conformidade com a INSPIRE;
- INSPIRE Geographic Name – tem a mesma função mas para o atributo Geographic-NamePropertyTypes.

3. Exportação para formato GML

Finalmente é possível exportar as transformações efectuadas no formato GML 3.2.1 conforme requerido na especificação de dados. Durante este processo é feita a validação do XML *schema*.

A visualização do resultado da transformação no formato GML 3.2.1 é

ilustrada na Figura 4. É possível observar que o código do GML obtido traduz as interrelações entre atributos tal como exigido pela Directiva.

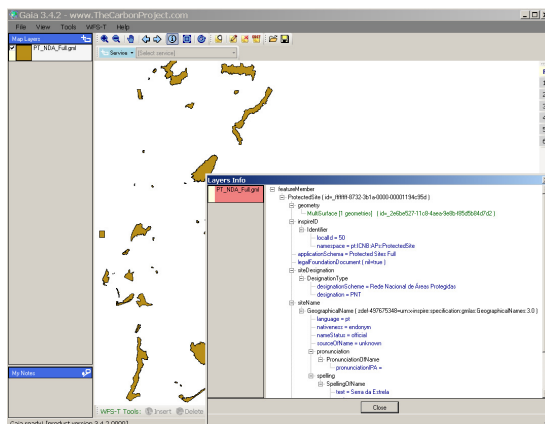


Figura 4. Exemplo dos dados transformados – Limites das Áreas Protegidas

A visualização do GML obtido não é ainda linear em determinados softwares (e.g. PostGIS, GeoServer, MapServer) que ainda não suportam a versão GML exigida pela Directiva.

A utilização do HALE demonstrou a sua adequabilidade para produzir de uma forma relativamente rápida e eficiente os dados harmonizados de acordo com as especificações de dados INSPIRE.

5. Conclusões e desenvolvimentos futuros

O projecto HUMBOLDT gerou contributos importantes para a aplicação da Directiva INSPIRE, nomeadamente através do desenvolvimento de ferramentas de harmonização de dados que permitem realizar a transformação necessária à obtenção de conjuntos de dados harmonizados. A disponibilização de ferramentas deste tipo é fundamental na actual fase de implementação da Directiva nos Estados-Membros.

O desenvolvimento deste caso de estudo demonstrou que o HALE é uma ferramenta promissora para a harmonização de conjuntos de dados geográficos porque permite o mapeamento e transformação de *schemas* e a sua transformação no formato GML 3.2.1 requerido pela Directiva.

Existem, no entanto, alguns obstáculos à completa aplicação das exigências da Directiva em termos de harmonização de conjuntos de dados geográficos, nomeadamente no que se refere à pouca consistência do formato GML e ao facto de este não se encontrar disponível em diversos softwares relevantes.

Um dos próximos passos será a disponibilização destes dados através do Serviço de Visualização INSPIRE. Pretende-se ainda produzir um manual de boas práticas para a harmonização de dados geográficos no domínio da Conservação da Natureza que possa servir de suporte às instituições públicas com responsabilidade na produção de dados no referido domínio.

Agradecimentos

Os autores agradecem a colaboração do ICNB concretizada nos contributos de Henrique Marinho, Ana Coito e Marta Marau da Unidade de Gestão de Sistemas do Instituto de Conservação da Natureza e Biodiversidade. Agradecem a Roderic Molina e Emanuele Rocatagliata da GISIG, parceiros no projecto HUMBOLDT envolvidos na aplicação das ferramentas de harmonização e no desenvolvimento dos Use Cases no contexto do cenário “*HS Protected Areas*”.

Referências

- [1] Jornal Oficial da União Europeia, Directiva 2007/2/CE do Parlamento Europeu e do Conselho, de 14 de Março de 2007, que estabelece uma infra-estrutura de informação geográfica na Comunidade Europeia (Inspire), <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:PT:PDF>.
- [2] Fonseca, A., Gomes, A.L., Furtado, D., Velasco, M., Belver, M.C., Martinez, A., O Papel das Infra-estruturas de Informação Geográfica no Apoio à Gestão Ambiental: Caso de estudo de dois Parques Naturais transfronteiriços. JIIDE 2010 – I Jornadas Ibéricas de Infra-estruturas de Dados Espaciais, Hotel Sana, Lisboa, 27-29 de Outubro, (2010).
- [3] HUMBOLDT project, <http://www.esdi-humboldt.eu/home.html>.
- [4] INSPIRE Data Specification on Protected Sites Guidelines v 3.1.0, <http://inspire.jrc.ec.europa.eu/>.
- [5] INSPIRE Generic Conceptual Model, (D2.5,Version 3.3), 2010, <http://inspire.jrc.ec.europa.eu/>.
- [6] Methodology for the development of data specifications: baseline version (D 2.6, Version 3.0), 2008, <http://inspire.jrc.ec.europa.eu/>.
- [7] NatureSDIplus – Best Network for SDI in Nature Conservation, <http://www.nature-sdi.eu/>.