

# Harmonização da Carta de Ocupação do Solo utilizando o HALE

**ZUNA, Teresa; FONSECA, Alexandra; FURTADO, Danilo; GOMES, Ana Luísa; SERRONHA, André; PATRÍCIO, Paulo**

A harmonização de dados geográficos é um dos principais desafios da implementação da Diretiva INSPIRE. Em 2015, os Grupos de Trabalho Temáticos (GTI-TE) criados em 2010, reunindo as autoridades públicas portuguesas responsáveis pela produção de conjuntos de dados geográficos (CDG) relacionados com os temas dos anexos da Diretiva foram reestruturados de acordo com os *clusters* Europeus, estando a ser promovida pela DGT uma nova dinâmica com o objetivo de se alcançar a conformidade dos CDG. Os 9 GTI-TE constituem-se como estruturas de colaboração e partilha para todos os assuntos relacionados com os conjuntos e serviços de dados geográficos produzidos por entidades públicas, com especial relevância para as questões de harmonização.

Este artigo analisa o processo de harmonização e os principais desafios surgidos no decorrer da sua aplicação prática à Cartografia de Ocupação do Solo (COS 2010). A COS é um dos CDG produzidos pela DGT com maior relevância para o desenvolvimento de estudos em Gestão e Planeamento Ambiental em Portugal.

O processo de harmonização incluiu diferentes fases começando pela análise dos standards e especificações da Diretiva INSPIRE para o tema II.2 *Land Cover*. Foi realizada uma revisão de metodologias e ferramentas que considerou os modelos de dados (UML), as ferramentas de transformação de dados espaciais (e.g. Hale, FME) e a sua aplicação. O processo de validação dos resultados de harmonização também foi analisado, considerando as metodologias preconizadas pela diretiva INSPIRE, o *Abstract Test Suit* (ATS) incluído nas especificações de dados, os tipos de testes a realizar (e.g. xsd; gml *schematron*; *thematic schematron*) e as ferramentas de validação disponíveis (e.g. XML Spy, Oxygen XML editor, eENVplus). A experiência de validação adquirida pela DGT em colaboração com a equipa de validação do eENVplus durante o projeto EAGLE 6 foi muito relevante para esta tarefa.

A aplicação dos requisitos de harmonização à COS implicou a criação de uma *matching table* com as correspondências entre os atributos da COS 2010 (*source schema*) e os atributos definidos nas especificações de dados do tema II.2 *Land Cover* (*target schema*). A ferramenta de transformação *HUMBOLDT Alignment Editor* (Hale) foi usada para produzir o GML de acordo com o INSPIRE. A validação dos resultados do processo de harmonização seguiu o ATS definido nas especificações de dados do tema *Land Cover*. Foram utilizadas diferentes ferramentas de validação, Oxygen e eENVplus validator, e os diferentes ficheiros de validação (LandCover.xsd, GML *schematron* e *thematic schematron*) para executar os diferentes tipos de testes. Os erros identificados e a sua correção foram devidamente documentados. Pretendeu-se obter uma versão do INSPIRE GML sem erros e contribuir para a criação de exemplos de boas práticas que visam apoiar o trabalho dos GTI-TE.

## **PALAVRAS-CHAVE**

INSPIRE, IIG, Harmonização de dados geográficos, HALE, Ocupação do Solo.

## **INTRODUÇÃO**

A DGT é a entidade responsável pela coordenação operacional do Sistema Nacional de Informação Geográfica (SNIG) e o Ponto de Contacto Nacional (PCN) para a Diretiva INSPIRE. É também responsável pela produção de cartografia onde se conta a produção de alguma cartografia temática, como é o caso da Carta de Ocupação do Solo (COS).

O SNIG, foi criado há mais de 20 anos, através do Decreto-Lei n.º 53/90, de 13 de fevereiro, tendo sido a primeira Infraestrutura de Informação Geográfica (IIG) desenvolvida na Europa e a primeira a ser disponibilizada na Internet, em 1995 [1]. Com base na rede de instituições do SNIG, o PCN INSPIRE criou em 2007 redes de instituições para a implementação da Diretiva INSPIRE. A Rede de Pontos Focais INSPIRE Core (RPF INSPIRE Core) abrange as autoridades públicas formalmente responsáveis pela produção dos conjuntos de dados geográficos (CDG) e serviços nacionais enquadrados nos temas enumerados nos três Anexos da Diretiva.

No contexto da RPF INSPIRE Core foram criados em 2010 e reestruturados em 2015 de acordo com os *clusters* europeus entretanto criados pela Comissão Europeia (CE), grupos de trabalho temáticos (GTI-TE) para clarificação das responsabilidades formais das instituições envolvidas, acompanhamento do trabalho desenvolvido nos *Clusters* Temáticos e estudo da aplicação das Disposições de Execução (DE) aos CDG e serviços de que são responsáveis.

Foi também criado um Grupo Transversal (GTI-TR) que visa apoiar de forma articulada as autoridades públicas na produção e publicação de metadados, no desenvolvimento de serviços de dados geográficos e na harmonização dos CDG, assim como na resposta às questões de âmbito Europeu surgidas nomeadamente no MIG-T e MIG-P.

O processo de harmonização, surge no âmbito da Diretiva INSPIRE como “o processo de desenvolvimento de um conjunto comum de especificações de dados, que viabilize o acesso aos dados geográficos através de Serviços, permitindo combinar dados de forma coerente”. Segundo o projeto HUMBOLDT (*Towards the Harmonisation of Spatial Information in Europe*) este processo é encarado como a “criação da possibilidade de combinar dados a partir de fontes heterogéneas, transformando-os em produtos com mínima ambiguidade, consistentes e integrados, sem preocupações para o utilizador final”.

A harmonização de dados geográficos é um dos principais desafios da implementação da Diretiva INSPIRE e surge como consequência das limitações existentes no acesso e utilização de dados geográficos aos diferentes níveis (e.g. do nível local ao europeu).

Entre outras questões relativas à integração de CDG podem enunciar-se a duplicação de informação, a incompatibilidade de formatos entre conjuntos e serviços de dados geográficos e os entraves à sua partilha e reutilização que contribuem para a falta de interoperabilidade técnica e semântica no acesso e utilização desses dados. Ao obrigar os Estados Membros (EM) a adotarem um conjunto de regras ou disposições de execução relativas às diferentes componentes das respetivas infraestruturas de informação geográfica (e.g. metadados, conjuntos de dados geográficos e serviços) a Diretiva INSPIRE pretende contribuir para ultrapassar os problemas de integração de dados geográficos anteriormente referidos.

Na prática, a harmonização é o processo que permite transformar os dados de origem num modelo de dados previamente descrito numa Especificação de Dados que define um modelo e um catálogo de objetos. Os modelos relativos a cada tema, definidos pela Diretiva INSPIRE, são uma forte recomendação para que todos os Estados Membros adotem os mesmos atributos, as mesmas relações entre informação e os mesmos domínios, tornando a informação geográfica interoperável ao longo da Europa.

Considerado como um dos aspetos mais críticos no contexto da aplicação da Diretiva INSPIRE, a harmonização de dados geográficos requer um esforço significativo por parte das instituições produtoras de dados geográficos dos diferentes EM que necessitam de apoio em termos técnicos e organizacionais para cumprirem este objetivo.

O conhecimento e experiência adquiridos pela DGT há já alguns anos através da participação da em projetos europeus no domínio das infraestruturas de informação geográfica especialmente focados na harmonização de dados geográficos requerida pela Diretiva INSPIRE (e.g. HUMBOLDT, GIS4EU, NatureSDIplus, HELM, eENVplus, EAGLE 6) constituem uma mais valia no apoio que a DGT pretende disponibilizar às instituições portuguesas com responsabilidades perante a Diretiva, envolvidas nos GTI-TE.

A participação da DGT no projeto EAGLE 6, concluído em setembro de 2015, focado no estudo e implementação de uma metodologia de harmonização da cartografia *CORINE Land Cover* e *Urban Atlas* em conformidade com as disposições de execução INSPIRE para a temática de ocupação do solo, surge com grande relevância em termos de capacitação para a harmonização de dados geográficos nas suas diversas fases (e.g. análise, mapeamento, transformação, validação) e considerando as características de ocupação de solo presentes em amostras de diferentes países. A DGT foi responsável pela tarefa de validação e consciente da importância da colaboração entre parceiros, projetos e países na partilha de conhecimento e experiências neste domínio, contou com a colaboração da equipa de validação do projeto eENVplus, que desta forma prestou um apoio essencial no âmbito do desenvolvimento do trabalho de validação do EAGLE 6. Um dos resultados deste projeto foi a produção do ficheiro *schematron Land Cover v.4*, que passou a estar também disponível no validador do eENVplus.

Sendo a Carta de Ocupação do Solo de Portugal Continental (COS) um dos CDG produzidos pela DGT com maior relevância para o desenvolvimento de estudos em gestão e planeamento ambiental em Portugal e na sequência da participação da DGT no projeto EAGLE 6, foi decidido avançar para a harmonização da cartografia de ocupação e uso do solo de Portugal Continental de acordo com as especificações da Diretiva INSPIRE.

Este artigo descreve o processo de harmonização e os principais desafios surgidos no decorrer da sua aplicação prática à Cartografia de Ocupação do Solo (COS 2010), recorrendo à utilização de uma ferramenta *open-source*, o *HUMBOLDT Alignment Editor* (Hale) na fase de transformação dos CDG.

## **DIRETIVA INSPIRE: HARMONIZAÇÃO DE CONJUNTOS DE DADOS GEOGRÁFICOS (CGD)**

A Diretiva INSPIRE do Parlamento Europeu e do Conselho, aprovada em março de 2007 (Diretiva 2007/2/EC), estabelece o enquadramento legal para a criação gradual de uma infraestrutura europeia de informação geográfica (IIG) de suporte às políticas ambientais ou atividades que tenham impactos no meio ambiente. Define que os Estados-membros da União Europeia são obrigados a gerirem e a disponibilizarem os dados e os serviços de informação geográfica de acordo com os princípios e as especificações técnicas de interoperabilidade de dados e serviços, utilização de serviços, princípios de acesso, partilha de dados e metadados definidos na Diretiva. Foi transposta para a legislação portuguesa com a publicação do Decreto-Lei n.º 180/2009, de 7 de agosto recentemente alterado pelo Decreto Lei n.º 84/2015, de 21 de maio.

A Diretiva abrange informação geográfica referente a um conjunto de temas divididos por 3 anexos. No que diz respeito à Cartografia de Ocupação do Solo (COS) esta encontra-se inserida nos temas II.2 Ocupação do Solo e III.4 Uso do Solo.

A harmonização de dados é descrita como o processo de desenvolvimento de um conjunto comum de especificações para conjuntos de dados geográficos, de forma a viabilizar o acesso a esses dados através de serviços, que permita combinar esses dados com outros dados harmonizados de forma coerente [2]. A harmonização envolve a transformação dos dados de origem (*source schema*) num modelo de dados descrito pelas especificações da Diretiva (*target schema*), em formato aberto e orientado para os serviços.

A interoperabilidade dos conjuntos e serviços de dados geográficos é um dos objetivos principais da Diretiva. Neste contexto, a interoperabilidade permitirá, sem qualquer intervenção manual, a partilha, combinação e interação com os serviços, [3]. O conceito de interoperabilidade é entendido como essencial na partilha de dados geográficos através de *webservices* garantindo a interação e a comunicação entre diversas fontes de CDG, permitindo desta forma a utilização, de forma coerente, dos dados geográficos [4].

O processo de harmonização envolve a análise dos modelos de dados, o preenchimento da *matching table* ou tabela de correspondências, a transformação do CDG para o *target schema*, a validação e a publicação, através de serviços do CDG (Figura 1).

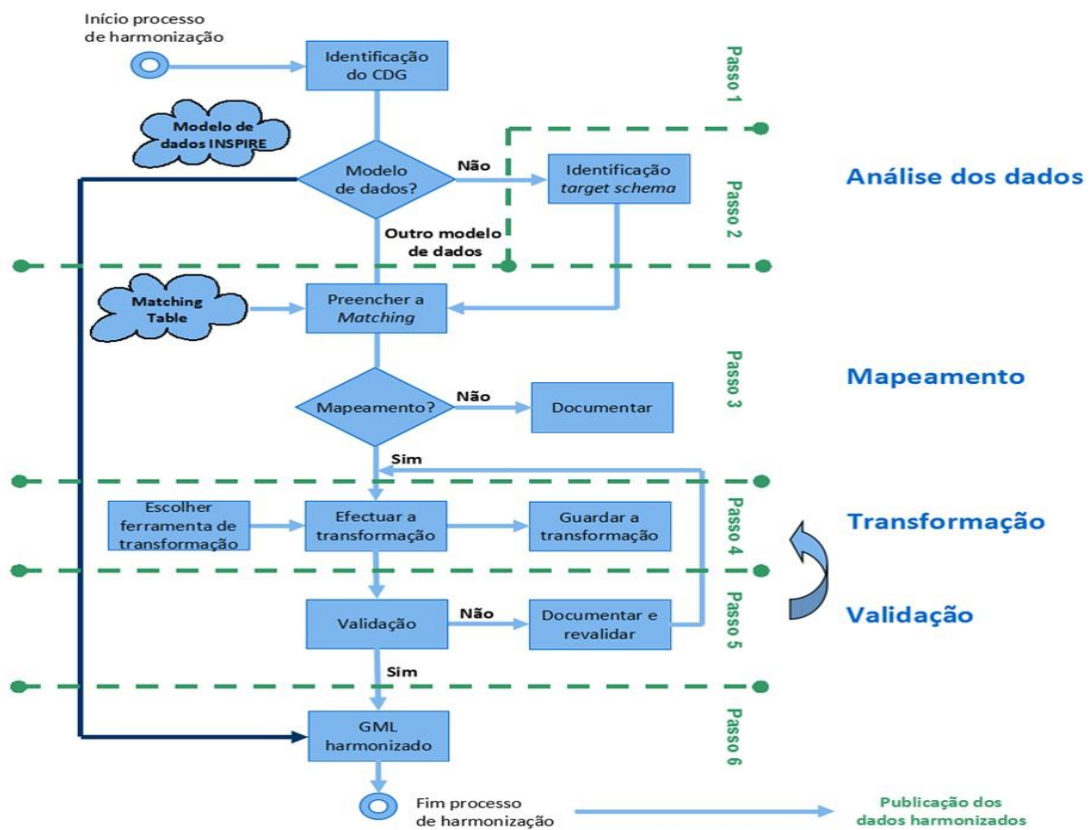


Figura 1: Fluxo de procedimentos no processo de harmonização do CDG (Fonte: [5]).

A análise dos modelos de dados passa pela compreensão dos dados originais e do *application schema* INSPIRE referente ao tema a que o CDG diz respeito. Identifica-se o formato, os atributos, a representação espacial, o sistema de coordenadas em que o CDG se encontra e a qualidade dos dados em geral, entre outras características. Este processo envolve ainda a compreensão do *target data model*, a interpretação do documento *Data Specifications* INSPIRE referente ao tema e os modelos de dados em linguagem UML.

As tabelas de correspondência, designadas por *matching tables* são utilizadas para estabelecer correspondência entre os atributos contidos no modelo de dados de origem (*source schema*) e a estrutura do modelo dos dados de destino (*target schema*). A *matching table* identifica e descreve as classes, os atributos, as enumerações e listas de códigos e associações entre as classes de ambos os modelos (Figura 2). É ainda utilizada para documentar o processo de harmonização, através do preenchimento de diversos campos que a constituem e adicionais onde são registados os aspetos relacionados com o estado do processo de harmonização.

Aplicativo Servico 1 de Coruña (Servico 1.0)						Aplicativo Servico 2 de Coruña (Servico 2.0)					
Feature type	Feature type description	Feature type definition	Geometry	Usage	Is mandatory	Category	Category definition	Category definition	Category definition	Usage	Is mandatory
Feature type	Feature type description	Feature type definition	Geometry	Usage	Is mandatory	Category	Category definition	Category definition	Category definition	Usage	Is mandatory
LandCover	LandCover	LandCover	MultiPolygon	LandCover	Yes	LandCover	LandCover	LandCover	LandCover	LandCover	Yes
LandCover	LandCover	LandCover	MultiPolygon	LandCover	Yes	LandCover	LandCover	LandCover	LandCover	LandCover	Yes
LandCover	LandCover	LandCover	MultiPolygon	LandCover	Yes	LandCover	LandCover	LandCover	LandCover	LandCover	Yes
LandCover	LandCover	LandCover	MultiPolygon	LandCover	Yes	LandCover	LandCover	LandCover	LandCover	LandCover	Yes
LandCover	LandCover	LandCover	MultiPolygon	LandCover	Yes	LandCover	LandCover	LandCover	LandCover	LandCover	Yes

Modelo Inspire  
LandCoverVector.xsd

Carta de Ocupação do Solo 2010 N5

Figura 2: Matching table (Fonte:[6]).

A transformação é o processo que se segue posteriormente à criação das correspondências entre os modelos de dados. Utilizando ferramentas de edição, formatação e conversão de dados é executado o processo de transformação, que, de acordo com as regras de implementação da Diretiva INSPIRE, irá dar origem a um ficheiro GML. Este processo irá transformar os dados de origem, independentemente do formato, num GML versão 3.2.1 com as características que foram definidas nas especificações de dados.

A validação é o processo de verificação da coerência apresentada pelo ficheiro GML relativamente à estrutura do *target schema* identificada de acordo com o tema do INSPIRE.

O objetivo final do processo de harmonização é a publicação dos CDG em serviços. A Diretiva INSPIRE define como serviços prioritários, o WMS (*Web Map Service*) e o WFS (*Web Feature Service*).

### O TEMA OCUPAÇÃO DO SOLO

A Carta de Ocupação do Solo de Portugal Continental para 2010 (COS2010) é uma cartografia temática que pretende caracterizar com grande detalhe a ocupação do solo no território de Portugal Continental. É composta por polígonos que representam unidades de ocupação homogéneas. Neste contexto, entende-se por unidade de ocupação do solo qualquer área de terreno superior ou igual 1 hectare, unidade mínima cartográfica (UMC), com distância entre linhas superior ou igual a 20 metros [7]. A COS2010 apresenta uma estrutura hierárquica com cinco níveis de detalhe, o que corresponde a 225 classes ao nível mais detalhado (N5). Esta nomenclatura é compatível com a nomenclatura da CORINE Land Cover aos primeiros três níveis.

O documento D2.8.II.2 *Data Specification on Land Cover - Technical Guidelines* fornece orientações para a aplicação das disposições previstas nas regras de implementação da Diretiva INSPIRE, para os CDG e serviços. Inclui ainda requisitos adicionais e recomendações consideradas relevantes para garantir ou melhorar a interoperabilidade dos CDG.

O modelo de dados proposto pela Diretiva INSPIRE para a representação em formato vetorial da ocupação do solo é, atualmente, o LandCoverVector.xml versão 4.0. O diagrama UML *Land Cover Vetor* descreve o modelo de dados para o tema da ocupação do solo e as relações existentes entre as várias classes presentes no modelo. O diagrama é constituído pelas seguintes classes: *Land Cover Dataset*, *Land Cover Observation*, *Land Cover Unit* e o *Land Cover Value*. O diagrama contém ainda o modelo *application schema Land Cover Nomenclature*, que diz respeito à nomenclatura correspondente ao CDG.

Para uma melhor compreensão do modelo de dados descreve-se, sucintamente, a estrutura do tema

da ocupação de solo. O *land cover dataset* consiste num conjunto de unidades de ocupação de solo, neste caso representadas por polígonos, e tem estabelecida uma ligação a uma *code list*. Esta é uma nomenclatura de classes de ocupação do solo, onde cada uma das classes é representada por um código e uma descrição. Para cada unidade de ocupação do solo, a superfície terrestre pode ser observada uma ou mais vezes. As várias observações permitem descrever as alterações ocorridas em termos de ocupação de solo, ao longo do tempo, estas são identificadas por um ou mais códigos constituintes da *code list*. É possível adicionar valores percentuais identificando a percentagem de ocupação de solo por cada classe em cada unidade de ocupação de solo [4]. A figura seguinte apresenta esquematicamente o modelo de ocupação do solo, em que as caixas cinzentas representam valores *voidable*. Os valores *voidable* referem-se aos requisitos definidos no modelo de dados de origem que cuja presença nos atributos dos CDG a harmonizar seja inexistente. O modelo de dados INSPIRE prevê esta situação através da determinação dos motivos *voidable*, sendo estes *unpopulated* (a propriedade não faz parte do CDG mantido pelo produtor de dados), *unknown* (o valor é desconhecido) e *withheld* (a propriedade é confidencial).

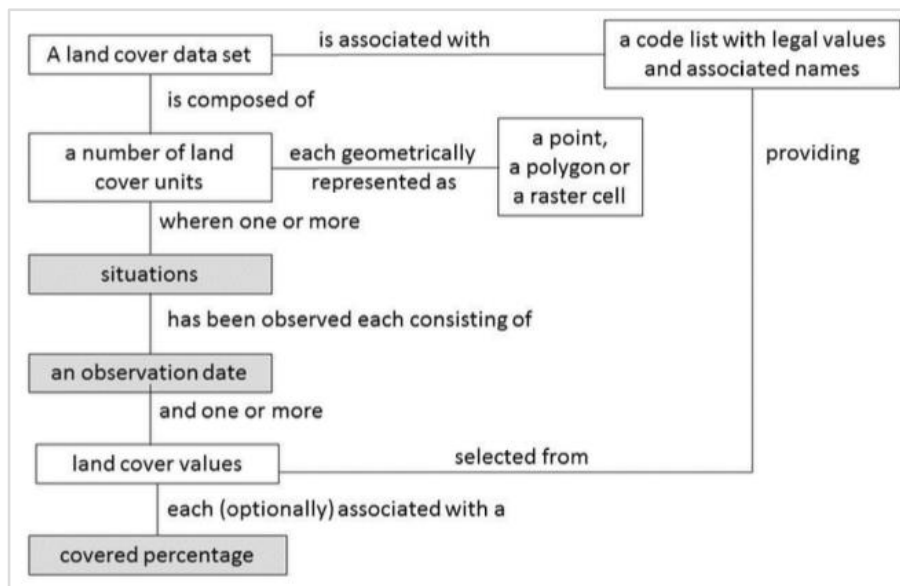


Figura 2: Modelo conceitual INSPIRE da ocupação do solo (Fonte [4]).

## HARMONIZAÇÃO DO CGD USANDO O HALE

O *HUMBOLDT Alignment Editor* (HALE) que atualmente já conta com a versão 2.9.4 foi desenvolvido através do projeto europeu HUMBOLDT ([www.esdi-humboldt.org](http://www.esdi-humboldt.org)) e pretende contribuir para a implementação da Diretiva INSPIRE. O HALE é uma ferramenta *open source*, desenvolvida com o objetivo de dar suporte e de facilitar processos de harmonização e transformação de CDG. Este software permite ao utilizador, estabelecer relações entre *schemas* (*source e target*) e transformar CDG, de forma automática, com base nas especificações definidas nos *application schemas*.

O HALE é bastante intuitivo e com capacidade de processar diferentes formatos de dados, fornecendo um bom suporte a linguagens do tipo XML/GML. Permite ainda visualizar o resultado do mapeamento durante o mesmo, através da validação online com o *application schema*.

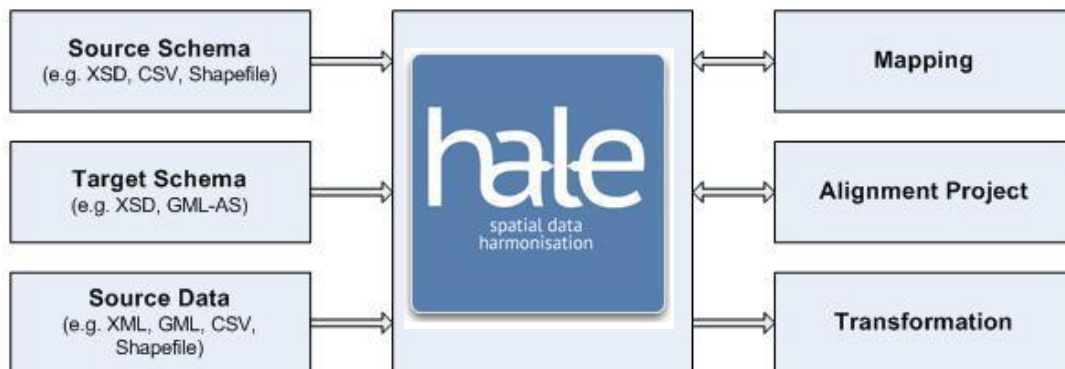


Figura 3: Input e output do HALE (Fonte:[8]).

O primeiro passo é importar a *source* e *target schemas* e a *source data* para o *HALE Schema Explorer*. Ao importar os CDG deve-se ter em conta a ISO 8859-11 definindo assim o alfabeto para latim. A importação do CDG de origem vai permitir a visualização gráfica, desde que o CDG apresente geometria e tenha associado um CRS (*Coordinate Reference System*). É necessário importar ainda as *code list* nos formatos .XML e .CSV.

O passo seguinte corresponde ao mapeamento entre as entidades do *source* e *target schemas* usando para isso a *matching table* mencionada anteriormente. Para estabelecer as relações de correspondência entre os *schemas* foram utilizadas as seguintes funções no HALE:

*Re-type* - Estabelece o mapeamento entre as *feature type* (do *source schema* para o *target schema*). Esta função indica que o *source* e o *target* são semanticamente iguais, i.e. para cada instância do *source type* vai ser criada uma instância no *target schema*. O mapeamento só pode ser estabelecido quando forem especificadas as relações de propriedade entre os *schemas*, só então é que podem estabelecer-se as relações de propriedade.

*Assign* - Atribui um valor a um atributo do *target schema*, quando não existe correspondência do lado do *source*.

*Formatted string* - Cria uma *string* com um formato baseado num padrão ou em variáveis de entrada (quando existe uma correspondência com o valor de entrada da função entre {} a *string* é validada, caso contrário o texto não é transformado).

*Rename* - Copia a propriedade de um atributo do *source schema* para um atributo do *target schema*.

*Classification* - Permite a utilização de *code lists* e outros domínios.

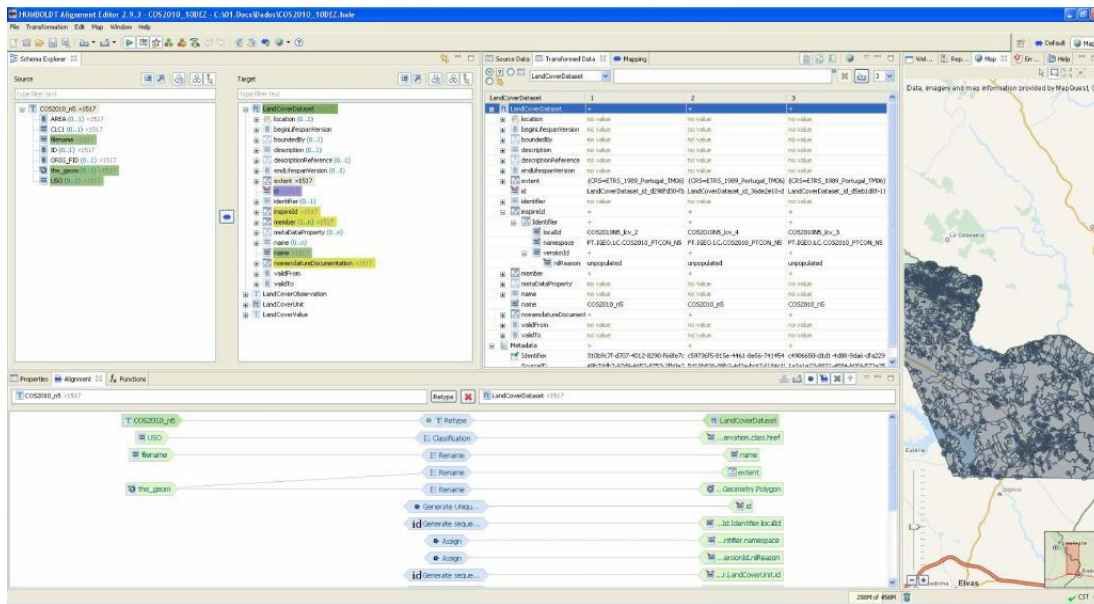


Figura 4: Interface do HALE durante o processo de harmonização da COS 2010 (Fonte: [5]).

Posteriormente ao estabelecimento de todas as relações foi possível exportar o resultado para o formato GML 3.2.1 conforme requerido pelas especificações de dados. O HALE permite a validação com o *Land Cover application schema* (XML).

O processo de validação do CDG (Figura 5) relativamente às especificações de dados da Diretiva INSPIRE requereu a utilização de ferramentas específicas. Neste caso houve a necessidade de utilizar um software/aplicação online de validação e edição de XML que permitiu validar e/ou identificar os erros existentes no ficheiro transformado.

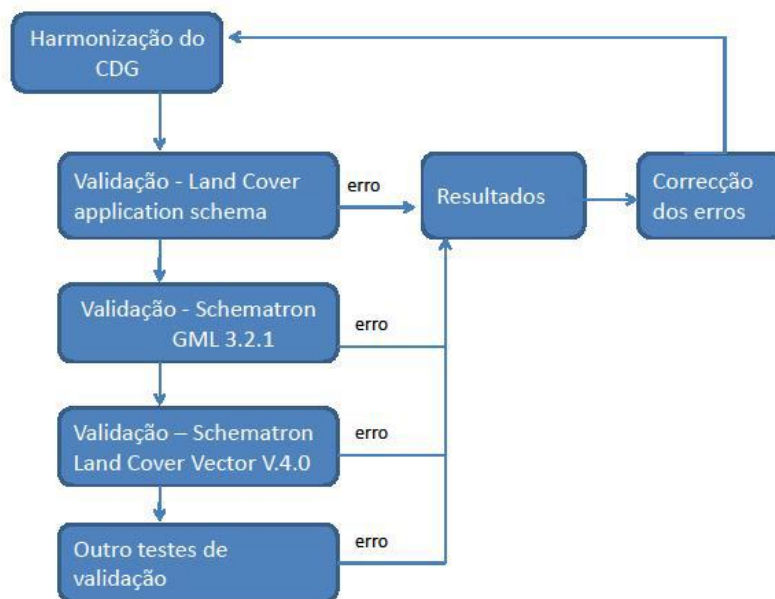


Figura 5: Metodologia do processo de validação (Fonte: adp. [9]).



A Comissão Europeia tem vindo a promover o financiamento de projetos de investigação, contribuindo para a aquisição e partilha de conhecimento no domínio da harmonização de CDG e serviços, apoiando o desenvolvimento de ferramentas essenciais no processo de harmonização de CDG. O projeto eENVplus [10] de cuja equipa a DGT fez parte, é um desses exemplos. O eENVplus disponibiliza um serviço de validação online e gratuito, ([http://cloud.epsilon-italia.it/eenvplus\\_new/ATS.htm?](http://cloud.epsilon-italia.it/eenvplus_new/ATS.htm?)), que permite a implementação dos ATS (*Abstract Test Suite*) incluídos no Anexo A das especificações de dados. Este conjunto de testes executável (ETS) verifica a conformidade dos conjuntos de dados GML em relação aos *application schemas* e também em relação à ISO 19136:2007 (*schematron* GML 3.2.1). Permite ainda a validação com ficheiros de tipo *schematron* para os temas já disponibilizados.

A DGT esteve envolvida em 2015 num projeto para a Agência Europeia do Ambiente, EAGLE 6, que procedeu à harmonização de dados do *CORINE Land Cover* e *Urban Atlas* de acordo com a Diretiva INSPIRE. Responsável pela tarefa de validação e consciente da importância da colaboração entre parceiros, projetos e países na partilha de conhecimento e experiência relativa a INSPIRE, a DGT assinou um *Memorandum of Understanding* (MoU) com a equipa de validação do projeto eENVplus, representada pela empresa EPSILON Italia, que desta forma prestou um apoio essencial no desenvolvimento do trabalho de validação realizado no âmbito do EAGLE 6. Um dos resultados do EAGLE 6 foi a produção do ficheiro *schematron Land Cover v.4*, disponível no validador do eENVplus.

O validador eENVplus foi utilizado como exemplo de execução do processo de validação automática do GML com uma ferramenta disponível online e gratuita. O GML resultante do processo de harmonização foi validado com o application schema *Land Cover Vector*, o *schematron* GML 3.2.1 e o *schematron Land Cover v.4*. A figura seguinte apresenta os resultados desse processo.

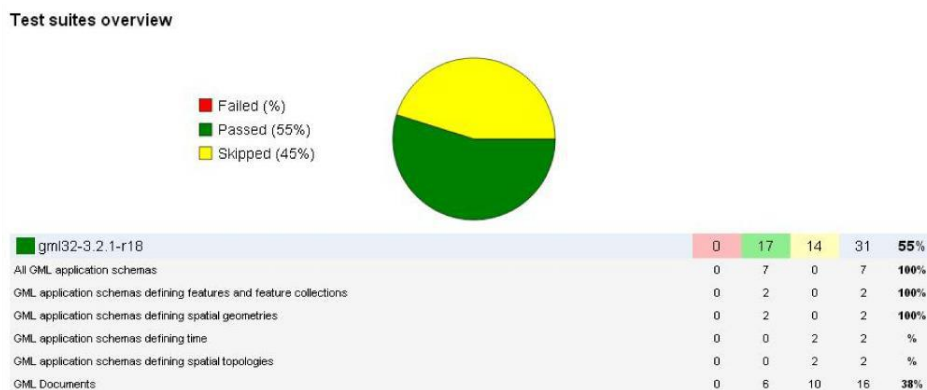


Figura 6: Resultado da sessão de validação com vista geral sobre os ATS realizados (Fonte: [6]).

Como auxiliar de validação do GML transformado foi ainda utilizado o software de edição de XML <oxygen/> XML Editor. Este é um software de proprietário que apresenta uma linha de aprendizagem relativamente pequena, sendo bastante intuitivo. O <oxygen/> permitiu a visualização do ficheiro GML em formato pretty XML e a validação com os esquemas de aplicação da Diretiva e com os *schematrons* GML 3.2.1. e *schematron Land Cover* versão 4. Através deste processo garantiu-se a conformidade da Carta de Ocupação do Solo 2010 com as especificações de dados INSPIRE.

A Figura 7 ilustra a utilização em SIG do ficheiro GML final da COS 2010 harmonizado de acordo com as especificações de dados INSPIRE para o tema *Land Cover*.

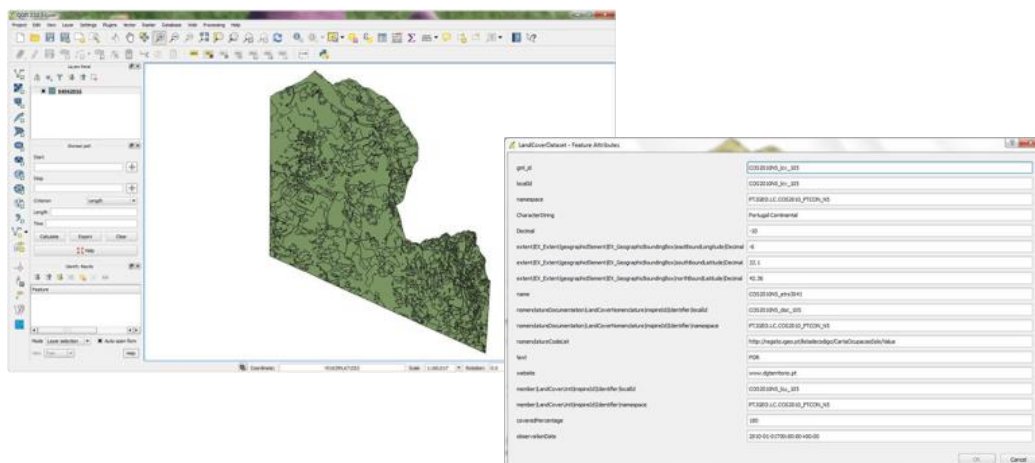


Figura 7: Utilização do ficheiro GML da COS 2010 no QGIS (Fonte: [6]).

## CONCLUSÕES

A interoperabilidade dos conjuntos e serviços de dados geográficos visa a partilha, combinação e interação de qualquer informação geográfica independentemente das fontes, facilitando processos de tomada de decisão aos vários níveis administrativos. A Comissão Europeia (CE) através da Diretiva INSPIRE pretende a criação de uma IIG comum aos Estados membros de forma a promover a partilha e o acesso à informação geográfica e desta forma promover oportunidades de utilização desta mesma plataforma.

Durante o processo de harmonização da COS2010 foram identificadas várias dificuldades, quer na transformação do CDG quer na validação do resultado com as especificações de dados INSPIRE.

A harmonização da COS2010 baseou-se numa primeira fase na identificação e estudo de projetos europeus que visaram a harmonização de CDG de acordo com o INSPIRE. Durante esta fase procedeu-se ainda à análise dos documentos referentes à Diretiva INSPIRE, o que requereu não só a leitura das DE gerais e específicas ao tema da ocupação do solo, como também a interpretação dos diagramas UML e dos *schemas* em linguagem XML, nomeadamente o *LandCoverVector*. A interpretação dos diagramas e dos *schemas* em XML requer um sólido conhecimento de linguagem UML e XML.

Os documentos referentes à Diretiva não só são extensos como se encontram interligados e a sua compreensão é um requisito para a boa execução dos trabalhos. Foi necessária a identificação e compreensão da matéria legislativa publicada até à data pela Comissão Europeia no âmbito do INSPIRE. A necessidade desta leitura deve-se ao facto do INSPIRE apresentar requisitos legais, técnicos e simples recomendações. De forma a que o processo de harmonização de dados seja bem sucedido, os requisitos legais e técnicos, bem como algumas recomendações têm de ser perentoriamente seguidos.

De forma a contribuir para a implementação a nível Europeu do INSPIRE a CE tem cofinanciado diversos projetos. O HALE, em formato *open source*, foi uma ferramenta criada nesse sentido, para dar suporte ao processo de harmonização de acordo com as DE do INSPIRE. Este software apresentou um bom desempenho durante o processo de transformação da COS2010 com as DE da ocupação do solo, tendo-se mostrado viável e com uma curva de aprendizagem relativamente pequena, encontrando-se ainda bem documentado. Para o processo de validação, também com origem num projeto da CE, o validador eENVplus demonstrou-se fácil de usar e com um bom desempenho. No entanto, para garantir a compreensão dos erros durante o processo de validação sentiu-se a necessidade de utilizar um software de proprietário, neste caso o *oXygen/XML editor*, com uma curva de aprendizagem relativamente pequena, através deste software foi possível identificar os erros e corrigi-los.

O processo de harmonização de um CDG exige um conhecimento abrangente ao nível dos dados de origem (*source data*). Sendo a Carta de Ocupação do Solo de Portugal Continental (COS) um dos CDG produzidos pela DGT mostrou-se essencial a experiência adquirida durante o processo de aquisição de informação geográfica e a análise à *posteriori* dessa informação que deu origem à COS2010. A boa compreensão dos dados de origem permitiu estabelecer relações credíveis (*schema mapping*) com o *target schema*.

O preenchimento da *matching table* foi essencial na documentação de todo o processo de harmonização. A *matching table* foi utilizada de forma a documentar todos os passos realizados durante o *schema mapping* tendo sido corrigida sempre que foi identificado um erro durante o processo de transformação e validação.

Durante o processo de transformação foram identificadas várias dificuldades que foram sendo ultrapassadas à medida que se aprofundava o conhecimento sobre as DE e os *application schemas*; o erro mais comum prendeu-se com o facto de por vezes ser difícil identificar com clareza o atributo do *target data* ao qual o atributo do *source* correspondia, estes erros foram facilmente corrigidos através da correta identificação do campo a preencher. Outra dificuldade prendeu-se com os atributos de característica *voidable* e a falta de informação por parte da *source data*. O *target schema* da ocupação do solo é bastante mais complexo e estruturado que a COS2010, logo durante o *schema mapping* não foi possível preencher todos os campos, tendo-se optado por utilizar a descrição *voidable*.

O processo de validação, iniciado no HALE através da validação do *application schema* permitiu a identificação e correção de alguns dos erros atempadamente. Para a validação com o *schematron* temático foi essencial a utilização do ficheiro desenvolvido pelo projeto EAGLE 6, que contou com o apoio da equipa de validação do projeto eENVplus.

O principal objetivo deste trabalho foi a aplicação das DE relativas ao tema II.2 *Land Cover da Diretiva INSPIRE* à carta de ocupação do solo de Portugal Continental de 2010, com o objetivo de produzir um ficheiro harmonizado de acordo com as regras estabelecidas pela Diretiva INSPIRE. Considera-se que este objetivo foi atingido uma vez que o resultado final do processo de harmonização foi um ficheiro GML válido. Pretendeu-se ainda descrever o processo de harmonização e demonstrar a sua aplicabilidade a outros CDG, com o intuito de poder servir de apoio às autoridades públicas que têm que proceder à harmonização dos CDG de que são responsáveis, suportando assim o papel da DGT como PCN para a Diretiva INSPIRE.

Este caso de estudo permitiu concluir a fiabilidade da harmonização de CDG de acordo com as especificações de dados do INSPIRE, contribuindo como exemplo de boas práticas no âmbito da harmonização de dados.

## REFERÊNCIAS

- [1] Masser: All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographic Information Science*, vol. 13, Issue 1 January 1999, pp (67-84) (1999).
- [2] INSPIRE: Definition of Annex Themes and Scope (D2.3). INSPIRE Data Specifications Drafting Team (2008).
- [3] European Parliament, Council of the European Union: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Official Journal of the European Union* (2007).
- [4] INSPIRE: Data Specification for the spatial data theme Land Cover (D2.8.II.2). INSPIRE Thematic Working Group Land cover (2013).
- [5] Zuna, T., Furtado, D.: Fluxo de procedimentos no processo de harmonização do CDG. In:

Workshop sobre harmonização de dados geográficos de acordo com as especificações INSPIRE, DGT, Lisboa (2015).

- [6] Zuna, T.: Harmonização de dados geográficos de acordo com a Diretiva INSPIRE. O caso da cartografia de Ocupação do Solo. DGT, Lisboa (2016).
- [7] DGT: Memória descritiva da Carta de Uso e Ocupação do Solo de Portugal Continental para 2010. DGT, Lisboa (2010).
- [8] Reitz, T.: HUMBOLDT Alignment Editor Manual 2.0.M2, (2010) [http://community.esdi-humboldt.eu/attachments/72/alignment\\_editor\\_manual\\_2010-03-31-M2.pdf](http://community.esdi-humboldt.eu/attachments/72/alignment_editor_manual_2010-03-31-M2.pdf)
- [9] Tracasa: Common Database on Designated Areas (CDDA) in conformity with the INSPIRE data specifications. Step4 - validation and conformance test. Tracasa, EEA (2014).
- [10] eENVplus - eEnvironmental services for advanced applications within INSPIRE, <http://www.eenvplus.eu>

## AUTORES

**Teresa ZUNA**  
*teresa.zuna@gmail.com*  
DGT

**Alexandra FONSECA**  
*afonseca@dgterritorio.pt*  
DGT

**Danilo FURTADO**  
*dfurtado@dgterritorio.pt*  
DGT

**Ana Luisa GOMES**  
*luisa.comes@dgterritorio.pt*  
DGT

**André SERRONHA**  
*aserronha@dgterritorio.pt*  
DGT

**Paulo PATRÍCIO**  
*ppatricio@dgterritorio.pt*  
DGT