

Ministério do Ambiente
Direção-Geral do Território



**Classificação da informação geográfica registada nas
Bases de Dados do SNIG**

Dezembro de 2017

Índice

1. Introdução	1
2. Objetivos	2
3. Atividades desenvolvidas	2
4. Conclusão	16

1. Introdução

O SNIG foi concebido em 1990, com a coordenação do Centro Nacional de Informação Geográfica (CNIG), como uma rede distribuída ligando entre si os produtores de informação georreferenciada, gráfica e alfanumérica, tendo passado a estar à disposição de todos os utilizadores através da Internet em maio de 1995, situação que colocou Portugal numa posição pioneira no domínio das Infraestruturas de Informação Geográficas (IIG).

A coordenação do SNIG é da responsabilidade da DGT, enquanto entidade que preside ao Concelho de Orientação (CO-SNIG) desta infraestrutura. Ao CO-SNIG compete a coordenação estratégica do SNIG. A DGT é também o ponto de contacto nacional para a aplicação da Diretiva INSPIRE em Portugal.

Em 2006 iniciou-se uma nova fase do SNIG com um modelo conceptual em conformidade com os princípios e normativos preconizados pela Diretiva INSPIRE.

Em 2014 foi realizada uma renovação do geoportal do SNIG, que incidiu fundamentalmente nos aspetos que se relacionam com as pesquisas no catálogo de metadados e no funcionamento do visualizador. Ainda em 2014, procedeu-se à criação do portal iGEO, que visa impulsionar a utilização de dados geográficos abertos e gratuitos pelo menos para a Administração Pública e academia, funcionando de modo complementar com a infraestrutura SNIG.

No início de 2015 a DGT iniciou uma nova abordagem em termos de SNIG e INSPIRE que se pretende colaborativa e participada para a criação da visão estratégica/plano de ação denominado SNIG2020. Esta abordagem colaborativa, permitiu a identificação de um conjunto de necessidades e melhoramentos a desenvolver no atual geoportal SNIG.

Neste trabalho pretende-se efetuar uma análise de todos os metadados existentes no SNIG, para identificar os que se encontram de acordo com os Documentos Técnicos do SNIG e atualizados e identificar os metadados que não estão de acordo com os Documentos Técnicos, ou que estejam desatualizados, identificando os problemas que apresentam, de modo a serem corrigidos pelos seus responsáveis e posteriormente proceder à sua migração ou em último caso, à sua eliminação.

2. Objetivos

Inventariar os metadados existentes que estão ou não de acordo com os Documentos Técnicos do SNIG:

- Metadados para pesquisa e interoperabilidade entre infraestruturas de informação geográfica;
- Perfil Nacional de Metadados de Informação Geográfica (Perfil MIG).

Inventariar os metadados existentes que se encontram desatualizados:

- Identificação dos contactos presentes nos metadados;
- Identificação dos nomes das entidades presentes nos metadados.

Compilação de listas com a identificação dos metadados que não estão preparados para efetuar a sua migração para o novo Geoportal SNIG:

- Especificação de alguns campos referentes a estes metadados;
- Solicitação da validação dos mesmos, por parte dos seus responsáveis;

Compilação de listas de código (code list) com os nomes harmonizados e normalizados das entidades responsáveis:

- Identificação nos metadados, de todas as variações dos nomes de cada entidade responsável pelo recurso, distribuição e metadados;
- Harmonização e normalização dos nomes das entidades responsáveis, que foram identificadas nos metadados, de acordo com o Sistema de Informação da Organização do Estado (SIOE);

Desenvolvimento de automatismos para facilitar e reduzir o tempo na obtenção dos inventários e listagens.

- Pedidos HTTP GET;
- Pedidos CSW HTTP Post;
- Scripts de código em linguagem Python.

3. Atividades desenvolvidas

Começou por se efetuar uma análise geral, nomeadamente, determinar o número de metadados publicados por cada utilizador (gestor de dados de cada entidade), a data da última publicação e a entidade a que pertencem. Para tal, entrou-se na área de “Edição e Publicação de Metadados”, com as credenciais de administrador e na pesquisa foi filtrada a informação

por “Proprietário” e ordenada a “Data” por ordem crescente. Esta análise permite identificar que utilizadores existem, a entidade que representam como gestor de metadados, o volume de informação com que enriquecem o geoportal, se têm uma participação ativa ou não ativa no SNIG e se os metadados são recentes ou antigos, o que poderá se traduzir em informação atualizada ou desatualizada. De salientar que para saber a entidade a que pertence cada um dos utilizadores, é necessário abrir um metadado (definiu-se que seria o mais recente) e ver o campo “Nome da Organização” responsável pela criação do metadado.

O resultado desta análise é apresentado na Tabela 1.

Número de utilizadores que publicaram metadados no Geoportal SNIG	
Gestores de metadados	Nº de utilizadores
Ativo (1 ou mais metadados introduzidos)	87
Não ativa (0 metadados introduzidos)	188
Total	275

Tabela 1 – Número de Metadados Publicados no SNIG pelos Utilizadores

Conforme podemos constatar na Tabela 1, existe um total de 275 gestores de metadados registados no geoportal, em que apenas 87 desses utilizadores contribuíram com informação. Assim, podemos verificar que cerca de 2/3 dos utilizadores não contribuíram com qualquer metadado, pelo que não faz sentido as suas contas de utilizador existirem e sobrecarregar a IDE. É ainda possível identificar a existência de alguns utilizadores do tipo “teste”, utilizadores com mais de uma conta e a possibilidade de existirem utilizadores que já não exercem funções como gestores de metadados.

Seguidamente é apresentado na Tabela 2, outro pequeno resumo da informação constante no Anexo “Número e Data de Metadados Publicados”, nomeadamente, o ano e número de utilizadores em que publicaram o seu último metadado no SNIG.

Ano e número de utilizadores em que publicaram o seu último metadado	
Ano de publicação	Nº de utilizadores
2017	34
2016	9
2015	1
2014	5
2013	1
2012	3
2011	14
2010	14
2009	6

Tabela 2 – Ano e número de utilizadores em que publicaram o seu último metadado

Com base na Tabela 2, podemos verificar que 2017 foi o ano em que houve mais utilizadores a publicar o seu último metadado, existindo no entanto, um elevado número de utilizadores que efetuaram a sua última publicação há já vários anos, podendo isto significar que para além de haver muitos utilizadores que não participam há vários anos no geoportal, os seus metadados

poderão estar desatualizados, já não se encontrando válidos ou até mesmo serem considerados dados históricos.

Após esta análise, considerou-se importante validar todos os metadados publicados pelos vários utilizadores no SNIG. Deste modo, decidiu-se que o *download* e posterior validação dos metadados, deveriam ser efetuados por utilizador, pois deste modo, era mais fácil entrar em contacto com o gestor desses metadados e informá-lo sobre quais os metadados que apresentavam problemas ao nível da sua validação. Neste sentido, após o validador, os metadados em que existia interesse em extrair mais informação seriam os que chumbavam na validação, já que os restantes, nomeadamente, os que passariam positivamente, estariam prontos para o processo de migração para o novo Geoportal SNIG. Esta tarefa revelou-se bastante dispendiosa ao nível do tempo e esforço, uma vez que para validar os metadados no validador do SNIG, é necessário efetuar o *download* e a validação do metadado, individualmente. De facto, seria interessante e vantajoso existir no geoportal, a opção de efetuar o download de todos os metadados pretendidos pelo administrador em simultâneo e ainda a posterior validação dos mesmos também em simultâneo. Assim, considerou-se que deveríamos inicialmente efetuar apenas o download e posterior validação dos metadados do utilizador *administrador* pertence à DGT e depois far-se-ia o mesmo procedimento para os restantes utilizadores. Todos os metadados deste utilizador foram descarregados e validados individualmente e manualmente. Na Tabela 3 apresenta-se um pequeno resumo dos resultados obtidos.

Resultado da validação dos metadados do utilizador <i>administrador</i>	
Passou no validador	2176
Chumbou no validador	57
Total	2233

Tabela 3 – Número de metadados que passaram ou chumbaram no validador

Conforme constatamos na Tabela 3, o utilizador *administrador* contribui para o SNIG com 2233 metadados, dos quais 2176 passaram positivamente no validador, estando prontos para serem migrados para o novo Geoportal e 57 chumbaram no validador, pelo que necessitam de ser corrigidos antes da sua migração.

É importante referir que os dados ID, Título, Data e Nome da Entidade Responsável de cada um dos metadados, foram inicialmente extraídos manualmente, mas uma vez que o volume de fichas com informação é muito elevado, considerou-se importante desenvolver uma aplicação em linguagem Python, para automatizar este processo de análise e extração de informação. Contudo, esta aplicação desenvolvida será abordada neste documento, futuramente.

No decorrer do projeto, a validação de todos os metadados conforme inicialmente previsto, não foi concretizada, tendo-se apenas validado os metadados do utilizador *administrador*. De facto, existem metadados no SNIG que não estão conforme o Perfil MIG, devido a terem migrado de plataformas mais antigas e não terem passado pelo validador. Todos os restantes metadados, nomeadamente, aqueles que foram carregados e validados na atual plataforma estão válidos. Os problemas de validação aplicam-se aos metadados mais antigos. Deste modo, considerou-se que a validação e a avaliação aos metadados não deveria ser a direta (metadado

a metadado, por utilizador), mas sim de forma geral, a todos os metadados (data do 1º e último metadado criado, por entidade). Quando se fizer a migração para o novo geoportal, os metadados mais antigos terão que ser validados. Uma vez que a DGT trabalha com as entidades e não com os utilizadores, optou-se por agrupar os metadados por entidade. Para além disso, os metadados das entidades podem ser pesquisados diretamente, o que já não acontece com os metadados por utilizador.

Após reflexão sobre a melhor forma de avaliar os metadados, considerou-se interessante e importante a determinação para cada Entidade Responsável pelos Metadados, o número de metadados por Tipo de Recurso. Inicialmente, procedeu-se à criação de interrogações para cada uma das Entidades Responsáveis e submetê-las ao Servidor SNIG. Nos primeiros resultados obtidos, verificou-se que o número de metadados que retornavam aos nossos pedidos, eram diferentes do número de metadados que deveriam retornar. Os motivos para este problema eram dois, nomeadamente:

- A especificação OGC do serviço CSW - OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile, do Open Geospatial Consortium Inc., anteriormente referida, não tem qualquer parâmetro que permita filtrar os metadados apenas pela Entidade Responsável pelo Metadado;
- A inexistência de uma Lista de Códigos (Code List) harmonizada e normalizada, com o nome das várias Entidades produtoras de metadados e consequentemente o aparecimento do nome das Entidades escrito de diversas formas.

Deste modo, a solução encontrada e definida foi a introdução e utilização de novos parâmetros e filtros no código Python, anteriormente criado. Assim, foi necessário determinar para cada uma das Entidades, todas as variações de nome existentes, pelo que foi desenvolvido um ficheiro de código Python, onde é solicitado para além dos ID, Título, Data e Nome da Entidade Responsável pelo Metadado, o Nome do Responsável pelo Recurso e Nome da Entidade Responsável pela Distribuição. Apesar de pretendermos obter a informação das Entidades Responsáveis pelos Metadados foram extraídos também os nomes de todas as outras Entidades presentes nos contactos. Pretende-se criar com esta extração uma Lista de Códigos para o nome das Entidades, assim como corrigir o nome de todas as Entidades presentes nos metadados. Do ficheiro resultante deste código, foram extraídos todos os Nomes das Entidades, para um outro ficheiro “.csv”, identificaram-se as Entidades que existem e para cada uma delas, todas as variações de nomes existentes nos metadados. Estas variações foram agrupadas por Entidade e com recurso ao site “SIOE - Sistema de Informação da Organização do Estado” (<https://www.sioe.dgaep.gov.pt/Default.aspx>) foram pesquisados, identificados e definidos os nomes harmonizados e normalizados, para cada uma das Entidades. Posteriormente, com base em todo este procedimento, foram então identificados e extraídos, para outro ficheiro “.csv”, os Nomes das Entidades Responsáveis pelos Metadados (harmonizados e normalizados) e as variações dos nomes existentes nos metadados.

Uma vez que possuía-mos esta informação, foi criado e desenvolvido um ficheiro de código Python genérico, para percorrer todos os metadados existentes no SNIG, os quais foram obtidos pelas nossas interrogações, submetidas ao servidor do SNIG e deste modo, extrair os

respetivos Nome da Entidade Normalizado e Harmonizado, Tipo de Recurso, ID, Título, Data, Entidade Responsável pelo Metadado, Email do Responsável pelo Metadado, Entidade Responsável pelo Recurso, Email do Responsável pelo Recurso, Palavras-chave e armazenar toda a informação extraída num ficheiro “.csv”.

Para uma melhor organização e arrumação dos dados, foi desenvolvido um outro ficheiro de código Python (ficheiro Python “Geração de Pastas e Cópia de Scripts Entidades” em Anexo), para percorrer o ficheiro “.csv” que contem os Nomes normalizados das Entidades Responsáveis pelos Metadados com as respetivas variações, criar uma pasta para cada uma das Entidades normalizadas, com o respetivo nome, e para dentro de cada uma delas, copiar o ficheiro de código Python genérico, já com o nome da respetiva Entidade normalizado.

Após correr o código para a criação, organização e arrumação das pastas, procedeu-se à edição dos ficheiros Python genéricos, que contêm as condições e filtros estabelecidos e que foram copiados para dentro da pasta de cada entidade. Em cada um dos ficheiros, especificamos as variações dos nomes das entidades, a serem localizados dentro dos metadados, bem como as diretorias das pastas das respetivas entidades, nas quais os ficheiros “.csv”, resultantes do código corrido, devem ser armazenados. Deste procedimento, obtiveram-se 218 pastas (uma pasta por cada Entidade Responsável pelos Metadados) e em cada uma delas, 2 ficheiros (1 ficheiro do tipo Python com o código para extração de dados e outro do tipo “.csv” com os dados extraídos pelo código). Esta organização tem vantagens não só ao nível da arrumação, como também, na extração dos dados das Entidades individualmente, não obrigando a extração dos dados das restantes Entidades. Contudo, se for necessário extrair novamente os dados de todas as Entidades, esta organização obriga a correr o código de cada uma das Entidades individualmente, ou seja, 218 vezes, o que se torna muito moroso e impede as pessoas de poderem realizar outras tarefas em simultâneo. Por este motivo, foi desenvolvido paralelamente um outro ficheiro de código Python, que basta correr apenas uma vez, e obtém-se um ficheiro “.csv” por cada Entidade (um total de 218 ficheiros), dentro da mesma pasta. Uma vez que estamos a falar em 218 Entidades, o processamento dos dados leva várias horas a ser concluído.

Com base nos 218 ficheiros obtidos, foram criadas e desenvolvidas duas tabelas, nomeadamente, “Entidade e Tipo de Metadado” e “Entidade e Ano do Metadado”. A primeira tabela, que é apresentada no ficheiro “Entidade e Tipo de Metadado” em Anexo, contem o Nome Normalizado das Entidades Responsáveis pelos Metadados e para cada uma delas, o número de Conjunto de Dados Geográficos, Serviços, Séries, Folhas, Conjunto de Dados Não Geográficos, Conjunto de Dados Não Conforme INSPIRE, Tipo de Entidade, o Total de Metadados e ainda o Total de Metadados de cada Tipo de Recurso de todas as Entidades. Esta informação é extremamente útil, pois permite-nos conhecer quais as Entidades com maior participação no Geoportal SNIG, o tipo de contributo, falhas no cumprimento das obrigações e normas estipuladas e ainda o panorama geral do Geoportal. A segunda tabela, apresentada no ficheiro “Entidade e Ano do Metadado” em Anexo, contem o Nome Normalizado das Entidades Responsáveis pelos Metadados e para cada uma delas, a Data do Primeiro Metadado Criado, a Data do Último Metadado Criado, o Tipo de Entidade e o Estado da Entidade. Esta informação é muito importante, para determinar quando é que cada uma das Entidades começou, deixou ou continua a participar e a contribuir para o Geoportal SNIG.

Terminada esta tarefa de avaliar e extrair todos os Nome da Entidade Normalizado e Harmonizado, Tipo de Recurso, ID, Título, Data, Entidade Responsável pelo Metadado, Email do Responsável pelo Metadado, Entidade Responsável pelo Recurso, Email do Responsável pelo Recurso e Palavras-chave, por entidade, iniciou-se uma nova tarefa, que consiste na realização de um conjunto de análises, segundo determinados critérios. Estes critérios são apresentados na Tabela 4, onde constam as interrogações (query) realizadas ao servidor do SNIG e o número de metadados que retornaram como resposta a essas interrogações, ou seja, os metadados que estão de acordo com os critérios por nós definidos.

Resultado global da análise de metadados filtrada por critérios	
Interrogações	Nº de metadados
Todos os metadados com as Palavras-chave "INSPIRECORE" e "iGEO" (intersecção de ambas)	177
Todos os metadados com as Palavras-chave "INSPIRECORE" ou "iGEO" (reunião de ambas)	1964
Todos os metadados que têm como Domínio CDG Prioritários e como Âmbito CDG identificados como prioritários, relacionados com Diretivas Europeias. (Palavra-chave começa por "Diretiva_"... ou "Directiva_"...)	157
Todos os metadados que têm como Domínio a Política de dados definida no Decreto-Lei nº 180/2009, de 7 de agosto, e que tem como Âmbito as Limitações ao acesso público, caso tal acesso possa prejudicar as relações internacionais, a segurança pública ou a defesa nacional. (Palavra-chave é "DL_180_2009_Art20_1")	0
Todos os metadados que têm como Domínio a Política de dados definida no Decreto-Lei nº 180/2009, de 7 de agosto, e que tem como Âmbito as Limitações ao acesso público (Palavra-chaves são "DL_180_2009_Art20_2a", "DL_180_2009_Art20_2b", "DL_180_2009_Art20_2c", "DL_180_2009_Art20_2d", "DL_180_2009_Art20_2e", "DL_180_2009_Art20_2f", "DL_180_2009_Art20_2g", "DL_180_2009_Art20_2h")	70
Todos os metadados que têm como Domínio a Política de dados definida no Decreto-Lei nº 180/2009, de 7 de agosto, e que tem como Âmbito as Condições de acesso a serviços: cobrança de taxas. (Palavra-chave é "DL_180_2009_Art18_2")	2
Todos os metadados com a Palavra-chave "INSPIRECORE"	1876
Todos os metadados com a Palavra-chave "iGEO"	265
Todos os metadados com a Palavra-chave "iGEO" e que o recurso seja do Tipo "Serviço"	244
Todos os metadados "sem restrições" (sem restrições de acesso e uso) e que o recurso seja do Tipo "Serviço"	602
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições" e que o recurso seja do Tipo "Serviço"	201
Todos os metadados com a Palavra-chave "iGEO" e que o recurso não seja do Tipo "Serviço"	21

Todos os metadados com a Palavra-chave "iGEO" e que o recurso seja do Tipo "Dataset"	21
Todos os metadados "sem restrições" (sem restrições de acesso e uso)	2415
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições"	215
Todos os metadados com a Palavra-chave "iGEO" e que não sejam "sem restrições" (com restrições)	50
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições" e que o recurso seja do Tipo "Serviço" e para o ano "2014"	9
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições" e que o recurso seja do Tipo "Serviço" e para o ano "2015"	5
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições" e que o recurso seja do Tipo "Serviço" e para o ano "2016"	40
Todos os metadados com a Palavra-chave "iGEO" e "sem restrições" e que o recurso seja do Tipo "Serviço" e para o ano "2017"	147
Todos os metadados em que o recurso seja do Tipo "Dataset"	4603
Todos os metadados em que o recurso seja do Tipo "Service"	1614
Todos os metadados em que o recurso seja do Tipo "Series"	1409
Todos os metadados em que o recurso seja do Tipo "Tile"	11416
Todos os metadados em que o recurso não seja do Tipo "Dataset", "Service", "Series" e "Tile"	54
Todos os metadados em que o recurso não seja do Tipo "Dataset", "Service", "Series", "Tile" e "NonGeographicDataset"	53
Todos os metadados em que o recurso seja do Tipo "NonGeographicDataset"	1
Todos os metadados em que o recurso seja do Tipo "Attribute"	0
Todos os metadados em que o recurso seja do Tipo "AttributeType"	0
Todos os metadados em que o recurso seja do Tipo "CollectionHardware"	0
Todos os metadados em que o recurso seja do Tipo "CollectionSession"	0
Todos os metadados em que o recurso seja do Tipo "DimensionGroup"	0
Todos os metadados em que o recurso seja do Tipo "Feature"	0
Todos os metadados em que o recurso seja do Tipo "FeatureType"	0
Todos os metadados em que o recurso seja do Tipo "PropertyType"	0
Todos os metadados em que o recurso seja do Tipo "FieldSession"	0
Todos os metadados em que o recurso seja do Tipo "Software"	0
Todos os metadados em que o recurso seja do Tipo "Model"	0
Todos os metadados em que o recurso seja um Serviço do Tipo "View" (WMS/WMTS)	1184
Todos os metadados em que o recurso seja um Serviço do Tipo "Download" (WFS/WCS)	422
Todos os metadados em que o recurso seja um Serviço do Tipo "Discovery" (CSW)	3
Todos os metadados em que o recurso seja um Serviço do Tipo	0

"Transformation" (WPS)	
Todos os metadados em que o recurso seja um Serviço do Tipo "Invoke"	0
Todos os metadados em que o recurso seja um Serviço do Tipo "Other"	1
Todos os metadados sobre a Carta de Ocupação do Solo (COS)	38
Todos os metadados da DGT e com a palavra-chave INSPIRECORE	81
Todos os metadados que existem no SNIG	19050

Tabela 4 – Interrogações e número de metadados resultantes

Na Tabela 4 é possível ver que para algumas interrogações existem vários metadados que respondem aos critérios definidos. Contudo, existem também várias interrogações em que o resultado não tem qualquer metadado, mas que se consideraram importantes em serem analisados.

Algumas das interrogações realizadas permitiram determinar a importância e utilidade de alguns Documentos Técnicos do SNIG e a adesão por parte dos utilizadores às especificações neles definidos. Um dos documentos é o “Perfil MIG” e o outro é o “Metadados para pesquisa e interoperabilidade entre infraestruturas de informação geográfica”, onde são definidas algumas palavras-chave de enorme importância, distribuídas em dois grupos. O primeiro grupo é apresentado na Tabela 5 e refere-se a palavras-chave que condicionam a monitorização anual realizada por Portugal para a CE e o funcionamento do portal iGEO.

Domínio	Palavra-chave	Âmbito
Interoperabilidade entre infraestruturas de informação geográfica	INSPIRECORE	CSDG de âmbito INSPIRE. Base para o filtro para o Geoportal INSPIRE e monitorização automática baseada nos metadados.
	iGEO	SDG para o Portal iGEO. Articulação entre o Portal iGEO e o catálogo do SNIG.
	DGT APA ...	Sigla da entidade

Tabela 5 – Palavras-chave que condicionam a monitorização anual realizada por Portugal para a CE e o funcionamento do portal iGEO

Se analisarmos as palavras-chave da Tabela 5 e que são utilizadas nas interrogações apresentadas na Tabela 4, podemos verificar pelos resultados obtidos no número de metadados que respondem a cada query, que houve uma boa afluência à utilização destas palavras, para a monitorização anual realizada por Portugal para a CE e o funcionamento do portal iGEO. Contudo, ao observarmos a Tabela 4, é possível identificar que existem problemas em alguns metadados, que utilizam de forma incorreta a palavra-chave “iGEO”. Esta palavra só pode ser adicionada em metadados do tipo Serviço e para articulação entre o Portal iGEO e o catálogo do SNIG, pelo que não podem existir restrições para estes recursos. Por exemplo, segundo os resultados das interrogações, existem:

- 21 metadados errados pois a palavra-chave “iGEO”, só pode ser utilizada em metadados do tipo Serviço:
 - Metadados com a palavra-chave “iGEO” = 265;

- Metadados com a palavra-chave “iGEO” e do tipo Serviço = 244;
 - 265 – 244 = 21 metadados errados.
- Os 21 metadados errados e que têm que ser corrigidos são do tipo Dataset, pois a palavra-chave “iGEO” não pode ser adicionada a metadados que não sejam do tipo Serviço:
 - Metadados com a palavra-chave “iGEO” e que não sejam do tipo Serviço = 21;
 - Metadados com a palavra-chave “iGEO” e que sejam do tipo Dataset = 21;
 - 21 metadados errados.
- Existem 43 metadados do tipo Serviço que estão incompletos, pois necessitam que lhes seja adicionada a palavra-chave “Sem Restrições”, já que a utilização da palavra-chave “iGEO” só pode ser adicionada a um metadado se este for do tipo Serviço e se for “Sem Restrições”:
 - Metadados com a palavra-chave “iGEO” e do tipo Serviço = 244;
 - Metadados com as palavras-chave “iGEO”, “Sem Restrições” e que seja do tipo Serviço = 201;
 - 244 – 201 = 43 metadados incompletos

Este exemplo mostra que de facto existem metadados criados com erros ou incompletos, necessitando de retificação por parte dos respetivos gestores de metadados. Este é um dos motivos pelos quais é extremamente importante a realização de interrogações mais complexas e relacionais ao servidor e a identificação da Entidade, ID, Título, Data do metadado e Palavras-chave de forma a detetar os problemas existentes no SNIG e posteriormente, informar de forma simples e direta cada um dos gestores de metadados sobre as alterações que têm que produzir nos seus metadados.

O segundo grupo é apresentado na Tabela 6 e refere-se a outras palavras-chave que contribuem para uma pesquisa mais eficiente dos metadados de informação geográfica.

Domínio	Palavra-chave	Âmbito
CDG prioritários	Diretiva_60_CE_2007, Diretiva_50_CE_2008, ...	CDG identificados como prioritários, relacionados com Diretivas Europeias.
Política de dados definida no Decreto-Lei nº 180/2009, de 7 de agosto	DL_180_2009_Art20_1	Limitações ao acesso público, caso tal acesso possa prejudicar as relações internacionais, a segurança pública ou a defesa nacional.
	DL_180_2009_Art20_2a, DL_180_2009_Art20_2b, DL_180_2009_Art20_2c, DL_180_2009_Art20_2d,	Limitações ao acesso público.

	DL_180_2009_Art20_2e, DL_180_2009_Art20_2f, DL_180_2009_Art20_2g, DL_180_2009_Art20_2h	
	DL_180_2009_Art18_2	Condições de acesso a serviços: cobrança de taxas.

Tabela 6 – Outras palavras-chave que contribuem para uma pesquisa mais eficiente dos metadados de informação geográfica

Se analisarmos as palavras-chave da Tabela 6 e que são utilizadas nas interrogações apresentadas na Tabela 4, podemos verificar que houve uma boa afluência à utilização destas palavras, contribuindo assim para uma pesquisa mais eficiente dos metadados de informação geográfica. Contudo, é de realçar que em determinados metadados a que se adicionaram as palavras-chave que se iniciam por “Diretiva_”, podem estar escritas com o antigo acordo ortográfico. Por esta razão, é necessário efetuar a interrogação com a palavra escrita no novo e no antigo acordo ortográfico, ou proceder à alteração do metadado, com a correção das palavras que se encontram com o antigo acordo.

A obtenção dos metadados segundo os critérios identificados na Tabela 4, não foi possível através da área de “Edição e Publicação de Metadados”, devido à escassez de filtros disponibilizados neste espaço. Deste modo, considerou-se necessário mudar a forma como se realizavam as análises e filtragem dos metadados, optando-se pelo método apresentado pela colega Sónia Fernandes, da Agência Portuguesa do Ambiente (APA), no “Blog inspire-pt” acessível no Geoportal do SNIG. O método consiste na utilização de pedidos HTTP GET, que permite realizar várias pesquisas aos metadados, com recurso a vários filtros, na mesma interrogação.

Neste sentido, foi necessário investigar e ler o documento da especificação OGC do serviço CSW - OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile, do Open Geospatial Consortium Inc., onde tem a explicação de todos os parâmetros e filtros que podem ser aplicados nas pesquisas aos metadados. Desta forma, desenvolveu-se em pedidos HTTP GET, todas as interrogações atrás mencionadas, sendo o código destas, apresentado no ficheiro “Código Especificacao OGC do servico CSW” em Anexo. Mas assim que se começou a obter os resultados para alguns destes pedidos, verificou-se que existia um conjunto de limitações, nomeadamente, limite da memória do Browser ao número de metadados que retornavam às nossas interrogações, bem como o facto de os resultados serem apresentados no ambiente web e não em ficheiros, como era pretendido.

Ao depararmo-nos com os problemas atrás referidos, optou-se por um novo método, nomeadamente, a utilização de pedidos CSW HTTP Post, para os quais a memória do Browser não coloca a restrição do número de metadados devolvidos e devolve os resultados em ficheiros, no formato “.xml”. Cada ficheiro devolvido pode conter um ou mais metadados, em função do número de metadados que respondam à interrogação submetida ao servidor. Este método que recorre ao software “CURL”, consiste inicialmente no desenvolvimento de ficheiros em formato “.xml”, com cada uma das interrogações realizadas individualmente, sendo o código de cada uma delas apresentado na pasta “Query Curl” em Anexo. Assim que

temos estes ficheiros finalizados, vamos correr cada um deles na “Linha de comandos”, com recurso ao código apresentado no ficheiro “Comandos CURL” em Anexo. Deste processo resultam então os ficheiros “.xml” com os metadados que respondem às nossas interrogações. É importante referir, que apesar de este método não ter as mesmas limitações que o método anterior, ao nível do número máximo de metadados devolvidos, este também tem uma limitação imposta pelo servidor. Por este motivo, em alguns casos, foi necessário criar certas interrogações partidas em mais de um ficheiro em função da limitação imposta pelo servidor.

Obtidos os ficheiros com os metadados que respondem às nossas interrogações, era necessário extrair os valores dos atributos Nome da Entidade Normalizado e Harmonizado, Tipo de Recurso, ID, Título, Data, Entidade Responsável pelo Metadado, Email do Responsável pelo Metadado, Entidade Responsável pelo Recurso, Email do Responsável pelo Recurso e Palavras-chave de cada metadado. Uma vez que estamos a falar de vários ficheiros com centenas e até mesmo milhares de metadados, a obtenção destes valores nos prazos estabelecidos, não era exequível manualmente. Desta forma, como anteriormente mencionado, considerou-se necessário desenvolver uma aplicação, em linguagem de “Python”, que percorresse todos os metadados à procura destes atributos e assim que os encontrasse, extraía os seus valores para um ficheiro em formato “.csv”, que era o pretendido pelo Eng^o. Henrique Silva e é também o formato mais utilizado neste tipo de trabalhos. O código desta aplicação para cada uma das interrogações é apresentado na pasta “Código para Extração de Metadados das Interrogações CURL” em Anexo. O desenvolvimento deste código foi realizado em várias etapas. Inicialmente era necessário encontrar um operador em “Python” que percorresse todos os metadados à procura dos atributos por nós pretendidos, neste caso “xml.dom.minidom.parse(filename_or_file, parser)”. Resolvida esta etapa, era necessário indicar os nós do ficheiro “.xml” em que se encontravam os atributos, tendo-se utilizado para esse objetivo o operador “getElementsByTagNameNS(namespaceURI,localName)”. Assim que encontrasse os atributos, extraía os seus valores. Pretendia-se que a extração destes valores fosse feita para um ficheiro “.csv”, em que os valores eram separados por “;” e que na primeira linha fossem escritos os nomes dos atributos. Ao fim destas etapas, começámos a correr a aplicação sobre os metadados e em alguns ficheiros ocorreram erros, nomeadamente:

- Acesso do código aos ficheiros de metadados negado;
- Nome para a mesma entidade escrito de diversas formas ou mudança de nome;
- Nome do Tipo de Recurso escrito de diversas formas;
- Existência de alguns metadados em que tinham Tags sem namespace;
- Atributos que eram pretendidos e que não estavam preenchidos;
- Nós de atributos que não estavam presentes nos metadados;
- Metadados que não tinham codificação “UTF-8”;
- Existência de mais do que um valor para o mesmo atributo.

O problema do acesso do código aos ficheiros de metadados negado, verifica-se quando se pretende aceder a um conjunto de ficheiros de metadados, bem identificados e armazenados numa pasta diferente da do código. Este problema não se verifica quando pretendemos que o código aceda apenas a um ficheiro de metadados, numa pasta diferente ou quando os ficheiros de metadados estão localizados na mesma pasta em que o código está armazenado. Na procura de uma solução, para este problema, muitas referências apontam para a necessidade de indicar no código a localização ou diretoria dos ficheiros de forma absoluta e não relativa. Esta hipótese foi testada, mas sem sucesso, pelo que se optou por uma outra solução, nomeadamente, copiar os ficheiros de metadados para a mesma pasta em que se encontra o código, e após o metadado ter sido percorrido e extraídos os seus valores, se procede à sua eliminação, de dentro desta pasta. Os motivos pelos quais estes ficheiros são eliminados, deve-se ao facto de terem uma dimensão muito grande (aproximadamente 319 MB) e de existirem várias análises, arrumadas e organizadas em pastas diferentes, pelo que levaria à existência de várias cópias em diversas pastas.

Quanto ao problema do nome para a mesma entidade estar escrito de diversas formas no metadado, por terem sido criados por diferentes utilizadores ou por mudança de nome (seja por alteração simples ou fusão de entidades), verificou-se que as análises realizadas em função do nome da entidade, retornavam com registos muito inferiores aos existentes e por vezes até mesmo nulos, quando sabíamos que na realidade existiam. Para contornar este problema, foi necessário começar por extrair todos os nomes das entidades escritos nos metadados (Nome da Entidade Responsável pelo: Recurso, Distribuição e Metadado). Após a extração, fez-se uma seleção de todas as variações dos nomes (não repetidos) e o seu agrupamento por entidade. Posteriormente, foram pesquisados e identificados os nomes harmonizados e normalizados para cada uma das entidades, com recurso ao site “SIOE - Sistema de Informação da Organização do Estado” (<https://www.sioe.dgaep.gov.pt/Default.aspx>), como mencionado anteriormente. Após esta identificação, foram inseridas no código variáveis com as variações dos nomes não repetidos das entidades, para cada entidade harmonizada e normalizada.

No que diz respeito ao problema do Tipo de Recurso escrito de diversas formas, verificou-se que dependendo do editor de metadados utilizado na criação dos metadados, o texto deste atributo no metadado, podia variar. A solução passou por identificar todas as variações de como estes atributos são escritos no metadado e inserir no código, variáveis para cada um dos Tipos de Recurso, com essas variações.

Em relação à existência de alguns metadados em que tinham Tags sem namespace, o código ao percorrer os metadados à procura das Tags por nós identificadas, não detetavam as Tags que não tinham o namespace identificado e que também eram por nós pretendidas. Assim, foi necessário identificar e acrescentar ao código os nós e atributos que tinham Tags sem namespace, de modo a detetá-los. Contudo, verificou-se que existiam várias combinações, que tinham que ser detetadas e identificadas no código, pois os resultados obtidos por este eram muitas vezes inferiores aos que existiam. Deste modo, foi necessário encontrar outra solução, pelo que se optou por utilizar um novo operador (`getElementsByTagNameNS`), que identifica as Tags por nós pretendidas, independentemente do seu namespace.

Quanto aos atributos que eram pretendidos e que não estavam preenchidos nos metadados, o código preenchia as células do ficheiro “.csv” com o valor do mesmo atributo, mas do metadado anterior, o qual era assumido por defeito e automaticamente. Desta forma, foi necessário indicar no código que quando o atributo não tivesse valor, a célula do ficheiro “.csv” teria que ser preenchida em branco ou vazia, evitando assim o seu preenchimento com valores errados.

Em relação aos nós de atributos que não estavam presentes nos metadados, o código ao percorrer os metadados à procura desses nós, gerava um erro. Deste modo, houve a necessidade de especificar que quando o nó do atributo não estivesse presente no metadado, este deveria ser ignorado, continuando a sua procura pelos restantes nós, preocupando-se apenas com os que estivessem presentes.

No que diz respeito aos metadados que não tinham codificação “UTF-8”, o código não conseguia interpretar todos os caracteres presentes nos metadados e conseqüentemente percorrê-los e extrair todos os dados disponíveis. Assim, foi necessário indicar que os metadados teriam que ser convertidos para a codificação “UTF-8”, com recurso à ferramenta “encoding='utf-8”.

Relativamente à existência de mais do que um valor para o mesmo atributo, verificou-se que o código ao extrair os valores destes atributos, apenas preenchia a célula do ficheiro “.csv” com o último valor identificado, não preenchendo também com os valores encontrados anteriormente. Desta forma, tivemos que indicar que sempre que fosse identificado mais do que um valor para o mesmo atributo, esse valor teria que ser inserido na mesma célula e concatenado ao valor encontrado anteriormente para esse atributo. Para além disso, indicámos também, que se fosse encontrado um valor igual ou repetido para o mesmo atributo, o mesmo seria ignorado, não sendo inserido na célula com a concatenação.

Esta aplicação foi desenvolvida para os seguintes cenários:

- Um ficheiro “.xml” com vários metadados dentro de si;
- Vários ficheiros “.xml” com vários metadados dentro de si;
- Uma pasta de vários ficheiros “.xml” com apenas um metadado dentro de si;
- Uma pasta de vários ficheiros “.xml” com vários metadados dentro de si.

Os resultados destas análises são apresentados na pasta “Resultado de Código para Extração de Metadados das Interrogações CURL” em Anexo, onde constam para cada uma das interrogações o ID, Título, Data e Nome da Entidade Responsável de cada metadado que responde à query submetida.

Uma vez que tínhamos para todas as interrogações os valores dos atributos atrás referidos, considerou-se importante e útil determinar para quais dos metadados do tipo Folha (Tile), existiam metadados do tipo Séries (Series). Deste modo foram cruzados manualmente os resultados destas duas interrogações, procurando pelo título das folhas nos títulos das séries. O resultado é apresentado no ficheiro “Tile Vs Series” em Anexo. Verificou-se que para cerca

de metade dos metadados do tipo Folha existem metadados do tipo Séries e que a outra metade não tem. Esta informação, poderá ajudar na identificação de quais os metadados do tipo Folha que poderão deixar de existir no geoportal, uma vez que já existem metadados do tipo Serviço, que abrangem todas essas folhas. Por outro lado, poderá também ajudar na determinação dos metadados do tipo Séries, que deverão ser criados, de forma a abranger todas as folhas, eliminando os seus metadados. Pretende-se com isto, efetuar uma limpeza dos metadados que existem em excesso no geoportal, tornando-o mais limpo, menos sobrecarregado, rápido e facilitar o processo de migração para o novo SNIG. Um dos critérios a ter em conta é o número de metadados do tipo folha que existe para uma série, pois se existir apenas um ou se for uma pequena parte de uma série, não faz sentido a substituição destes metadados pelo de uma série. Outro critério é a data em que foram efetuadas as folhas, pois por vezes são realizadas com um intervalo de tempo entre si grande, pelo que as características e o tipo de ocupação e uso do solo são bastante diferentes entre folhas, não sendo indicado para estes casos a criação de um metadado do tipo Série.

Em complemento deste trabalho, foram também realizadas as seguintes análises:

- Tipo de Recurso;
- Tipo de Entidade;
- Nº de Metadados por Tipo de Entidade;
- Tipo de Entidade por Estado da Entidade;
- Tipo de Recurso por Tipo de Entidade;
- Tipo de Entidade por Tempo Ativo;
- Tipo de Recurso por Entidade do Continente, Região Autónoma dos Açores (RAA) e Região Autónoma da Madeira (RAM);
- Entidades do Registo Nacional de Dados Geográficos por Tipo de Entidade;
- Tipo de Entidade por Número de Entidades e por Número de Metadados;
- Folhas Vs Séries por Número de Folhas e por Número de Séries;
- Número de Folhas por Série;

4. Conclusão

Esta classificação da informação geográfica incidiu sobre os metadados existente no Geoportal SNIG. Para esta classificação foram desenvolvidas ferramentas que permitiram realizar todas as análises pretendidas.

Com as ferramentas desenvolvidas será possível realizar novamente todas as análises efetuadas e outras que eventualmente vierem a ser consideradas necessárias. É de realçar que este trabalho acabou por permitir detetar lacunas e problemas existentes nos metadados, bem como facilitar a coordenação entre as várias entidades, no sentido de ultrapassar essas lacunas.