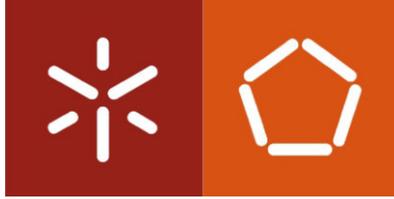




Universidade do Minho
Escola de Engenharia

Rafaela Oliveira da Cunha

**Análise da qualidade dos dados do European
Data Portal nas categorias de Governo e Setor
Público, Transportes e Educação, Cultura e
Desporto**



Universidade do Minho
Escola de Engenharia

Rafaela Oliveira da Cunha

Análise da qualidade dos dados do European Data Portal nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto

Dissertação de Mestrado
Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação

Trabalho efetuado sob a orientação da
Professora Doutora Ana Alice Rodrigues Pereira Baptista
e co-orientadora
Doutora Morgana Carneiro de Andrade

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição

CC BY

<https://creativecommons.org/licenses/by/4.0/>

AGRADECIMENTOS

A realização deste trabalho representa o fim de um ciclo na minha vida, o término de uma longa etapa, uma das maiores etapas e conquistas da minha vida.

O caminho para cá chegar não foi fácil, houve momentos em que corria tudo bem, mas também houve muitos momentos em que tudo parecia que ia desabar. Apesar disto, a vontade de vencer era mais forte, a capacidade de resiliência, o apoio da minha família e amigos e a fé em acreditar que tudo se resolveria, fazia com que todos os obstáculos fossem ultrapassados.

Assim, gostaria de agradecer aos meus pais por todo o apoio e ajuda até chegar aqui, sempre acreditaram em mim e nunca me deixaram desistir. Sempre fizeram tudo por mim e me ajudaram em tudo o que podiam e não há forma de agradecer por isso.

Queria também agradecer à minha orientadora, professora Doutora Ana Alice e à minha co-orientadora Doutora Morgana Andrade por toda a disponibilidade e apoio ao longo destes últimos meses.

Por último, gostaria de agradecer à Universidade do Minho por me acolher durante este percurso.

Obrigada por tudo.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio, nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

RESUMO

Análise da qualidade dos dados do European Data Portal nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto

Em alguns países, tem sido feito um esforço no sentido de disponibilizar, online e livre de encargos, dados do setor público, que, posteriormente, são publicados como dados abertos. Uma das iniciativas mundiais que agrega e disponibiliza dados abertos é o European Data Portal. Ele tem como finalidade recolher e disponibilizar dados e respetivos metadados do setor público que se encontram disponíveis em portais de dados abertos de vários países europeus. O presente estudo tem como objetivo determinar em que medida os conjuntos de dados das categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto publicados no Portal atendem às Boas Práticas para Dados na Web recomendadas pelo World Wide Web Consortium. A técnica de amostragem foi por conveniência. Com os conjuntos de dados ordenados pelo último modificado e filtrados pela classificação de excelente ou bom+, foram analisados 50 conjuntos de dados de cada categoria, desde que não fossem idênticos. A avaliação dos conjuntos de dados revelou que a categoria de Governo e Setor Público e Transportes apresentam conjuntos de dados com melhores classificações, tendo sido encontrados para ambos 11 conjuntos de dados com a classificação de suficiente e 39 com a classificação de bom. Na categoria de Educação, Cultura e Desporto, foram encontrados 34 conjuntos com a classificação de bom e 16 com a classificação de suficiente. Na análise dos conjuntos de dados, as Best Practices mais observadas foram: BP1, BP2, BP4, BP5, BP8H, BP10, BP11, BP12, BP13C, BP16, BP17, BP19, BP29, BP32 e BP34. E as menos observadas foram: BP3, BP6, BP7H, BP7C, BP8C, BP9, BP13H, BP14, BP15, BP21, BP22H, BP22C, BP30 e BP35. Esses resultados revelam que, em certas situações, a avaliação dos conjuntos de dados realizada pelo Portal corresponde à desenvolvida à luz das Boas Práticas para Dados na Web. Em outras situações, as avaliações são diferentes, por vezes quase opostas às do European Data Portal. Num trabalho futuro, pretende-se fazer o mesmo para as demais categorias presentes no Portal e elaborar recomendações específicas para os disponibilizadores de dados.

Palavras-chave: Dados Abertos; Data on the Web Best Practices; European Data Portal; Qualidade dos dados.

ABSTRACT

Analysis of the data quality of the European Data Portal in the categories Government and Public Sector, Transport and Education, Culture and Sport

In some countries, an effort has been made to make public sector data available online and free of charge, which is then published as open data. One of the worldwide initiatives that aggregates and makes open data available is the European Data Portal. It aims to collect and make available public sector data and related metadata that are available on open data portals in several European countries. This study aims to determine to what extent the datasets from the categories of Government and Public Sector, Transport and Education, Culture and Sport published on the Portal meet the Good Practices for Data on the Web recommended by the World Wide Web Consortium. The sampling technique was by convenience. With the datasets sorted by last modified and filtered by the rating of excellent or good+, 50 datasets from each category were analyzed, provided they were not identical. The evaluation of the datasets revealed that the category of Government and Public Sector and Transports present datasets with better ratings, with for both 11 datasets found with the rating of sufficient and 39 with the rating of good. In the Education, Culture and Sport category, 34 sets were found with the rating of good and 16 with the rating of sufficient. In the data set analysis, the most observed Best Practices were: BP1, BP2, BP4, BP5, BP8H, BP10 BP11, BP12, BP13C, BP16, BP17, BP19, BP29, BP32, and BP34. And the least observed were: BP3, BP6, BP7H, BP7C, BP8C, BP9, BP13H, BP14, BP15, BP21, BP22H, BP22C, BP30 and BP35.

These results show that in some situations, the evaluation of the datasets performed by the Portal corresponds to that developed in light of the Good Practices for Web Data. In other situations, the evaluations are different, sometimes almost opposite to those of the European Data Portal. In future work, it is intended to do the same for the other categories present in the Portal and to develop specific recommendations for data providers.

Keywords: Data on the Web Best Practices; Data quality; European Data Portal; Open Data.

ÍNDICE

| | |
|---|------|
| Agradecimentos | iii |
| Declaração de Integridade | iv |
| Resumo..... | v |
| Abstract | vi |
| Lista de Abreviaturas e Siglas | ix |
| Lista de Figuras..... | xi |
| Lista de Quadros e Tabelas..... | xiii |
| 1. Introdução | 1 |
| 1.1 Dados abertos..... | 1 |
| 1.2 Iniciativas | 3 |
| 1.3 Problema/Questões..... | 5 |
| 1.4 Objetivo Geral e Específicos | 7 |
| 1.4.1 Objetivos Específicos | 7 |
| 1.5 Estrutura da Dissertação | 8 |
| 2. Fundamentação Teórica..... | 9 |
| 2.1 Dados Abertos | 9 |
| 2.2 Dados Abertos Governamentais..... | 11 |
| 2.3 Melhores Práticas de Dados Abertos Governamentais | 13 |
| 2.4 European Data Portal | 14 |
| 2.5 Princípios FAIR | 19 |
| 2.6 Boas Práticas para os Dados na Web (DWBP)..... | 20 |
| 3. Enquadramento Metodológico | 32 |
| 3.1 Procedimentos Metodológicos | 33 |

| | |
|--|-----|
| 3.2 Estudo Exploratório | 35 |
| 3.3 Estudo Final | 38 |
| 3.3.1 Critérios para a Análise dos Conjuntos de Dados | 42 |
| 3.4 Procedimentos Adotados | 46 |
| 4. Resultados e Discussão | 50 |
| 4.1 Correspondência entre Indicadores EDP e Indicadores DWBP | 50 |
| 4.2 Análise do Portal: Estudo Exploratório | 53 |
| 4.3 Análise do Portal: estudo final | 57 |
| 5. Conclusões | 76 |
| Glossário | 78 |
| Bibliografia | 81 |
| Índice de Apêndices | 84 |
| Anexo 1- Estudo Exploratório | 86 |
| Anexo 1- Estudo Exploratório | 87 |
| Anexo 2- Estudo Final Governo e Setor Público | 89 |
| Anexo 2- Estudo Final Governo e Setor Público | 90 |
| Anexo 3- Estudo Final Transportes | 94 |
| Anexo 3- Estudo Final Transportes | 95 |
| Anexo 4- Estudo Final Educação, Cultura e Desporto | 99 |
| Anexo 5- Classificação Governo e Setor Público | 103 |
| Anexo 5- Classificação Governo e Setor Público | 104 |
| Anexo 6- Classificação Transportes | 105 |
| Anexo 6- Classificação Transportes | 106 |
| Anexo 7- Classificação Educação, Cultura e Desporto | 107 |
| Anexo 7- Classificação Educação, Cultura e Desporto | 108 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|--|
| APIs | Application Programming Interfaces |
| BP | Best Practice |
| CC | Creative Commons |
| CDO | Chief Data Officer |
| CSV | Comma-Separated-Values |
| DCAT-AP | Data Catalog Vocabulary – Application Profile |
| DOI | Digital Object Identifier System |
| DWBP | Data on the Web Best Practices |
| EDP | European Data Portal |
| EFTA | Associação Europeia de Comércio Livre |
| EISPs | Electronic Identifier Serial Publications |
| EUA | Estados Unidos da América |
| FAIR | Fairness, Accuracy, Interoperability and Reuse |
| HTML | HyperText Markup Language |
| IEINs | International eBook Identifier Numbers |
| INDA | Infraestrutura Nacional de Dados Abertos |
| ISNI | International Standard Name Identifier |
| ISO | International Organization for Standardization |
| JSON | JavaScript Object Notation |
| OCDE | Organização Europeia de Cooperação Económica |
| OGC | Open Geospatial Consortium |
| OGD | Open Government Data |
| PSI | Public Sector Information |
| PURL | Persistent Uniform Resource Locators |
| RDF | Resource Description Framework |
| REST | Representational State Transfer |
| VIAF | Virtual International Authority File |
| URN | Uniform Resource Name |

| | |
|------|---|
| WMO | World Meteorological Organization |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |
| XRIs | Identificadores de recursos extensíveis |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Pontos máximos para cada dimensão..... | 33 |
| Figura 2 – Escala de Pontos para cada classificação. | 35 |
| Figura 3 – Conjuntos de dados encontrados para a categoria Governo e Setor Público, Transportes, Educação, Cultura e Desporto..... | 35 |
| Figura 4 – Conjuntos de dados encontrados para a categoria de Governo e Setor Público, com a classificação de excelente e ordenado pela última alteração..... | 37 |
| Figura 5 – Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de excelente e ordenado pela última alteração. | 37 |
| Figura 6 – Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de Bom+ e ordenado pela última alteração..... | 37 |
| Figura 7 – Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de excelente e ordenado pela última alteração... | 38 |
| Figura 8 – Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Bom+ e ordenado pela última alteração. | 38 |
| Figura 9 – Conjuntos de dados encontrados para a categoria de governo e setor público, com a classificação de excelente e ordenado pela última alteração..... | 40 |
| Figura 10 – Conjuntos de dados encontrados para a categoria de governo e setor público, com a classificação de Bom+ e ordenado pela última alteração..... | 40 |
| Figura 11 – Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de excelente e ordenado pela última alteração. | 40 |
| Figura 12 – Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de Bom+ e ordenado pela última alteração..... | 41 |
| Figura 13 – Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Excelente e ordenado pela última alteração... | 41 |
| Figura 14 – Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Bom+ e ordenado pela última alteração. | 41 |
| Figura 15 – Número de conjuntos de dados da categoria de Governo e Setor Público em que cada BP está presente. | 54 |

| | |
|--|----|
| Figura 16 – Número de conjuntos de dados da categoria de Transportes em que cada BP está presente..... | 55 |
| Figura 17 – Número de conjuntos de dados da categoria de Educação Cultura e Desporto em que cada BP está presente. | 56 |
| Figura 18 – Número de conjuntos de dados da categoria de Governo e Setor Público em que cada BP está presente. | 59 |
| Figura 19 – Número de conjuntos de dados da categoria de Transporte em que cada BP está presente..... | 60 |
| Figura 20– Número de conjuntos de dados da categoria de Educação, Cultura e Desporto em que cada BP está presente. | 62 |
| Figura 21 – Número e percentagem de conjuntos de dados por classificação, na categoria de Governo e Setor Público..... | 68 |
| Figura 22 – Número e percentagem de conjuntos de dados por classificação, na categoria de Transportes..... | 71 |
| Figura 23– Número e percentagem de conjuntos de dados por classificação, na categoria de Educação, Cultura e Desporto..... | 74 |

LISTA DE QUADROS E TABELAS

| | |
|---|----|
| Tabela 1 – Pontuação de cada BP | 47 |
| Quadro 1 – Correspondência entre indicadores EDP e indicadores DWBP..... | 51 |
| Tabela 2 – Escala percentual para cada classificação | 65 |
| Tabela 3 – Classificação dos conjuntos de dados na categoria de Governo e Setor Público . | 65 |
| Tabela 4 – Classificação dos conjuntos de dados na categoria de Transportes..... | 68 |
| Tabela 5 – Classificação dos conjuntos de dados na categoria de Educação, Cultura e Desporto..... | 71 |

1. INTRODUÇÃO

1.1 Dados abertos

Quando se pretende referir o termo “dados” este pode variar notavelmente entre pesquisadores e, mais ainda, nas diferentes áreas de conhecimento. Esta diversidade em torno do conceito de dados, deve-se ao facto destes serem gerados para diferentes fins, por diferentes comunidades e por diferentes processos. Os dados podem-se referir, por exemplo, a números, imagens, textos, vídeos, áudio, *software*, algoritmos, modelos, simulações, entre outros (Sayão & Sales, 2014).

Alguns dados possuem valor imediato e durante muito tempo, enquanto outros vão adquirir valor com o decorrer do tempo.

Os dados podem ser distinguidos entre: dados observacionais, obtidos através de observações diretas, podendo estar associados a lugares ou tempo específico; dados computacionais, que se referem a modelos computacionais ou simulações; e dados experimentais, quando os dados são originados de situações laboratoriais, como por exemplo, medidas de uma reação química (Sayão & Sales, 2014).

Primeiramente é necessário ter em consideração que nos dias atuais, a grande quantidade de dados que se encontra online são recolhidos pelo setor público e posteriormente irão ser publicados como dados abertos (*Open Data*). Como dados abertos, entende-se toda “[...] informação que pode ser utilizada, modificada e partilhada livremente por qualquer pessoa e para qualquer fim” (Carrara, 2015, p. 3, tradução própria).

Dados abertos é um termo relativamente novo, com origem nas mesmas raízes que “*Open Source*” e “*Open Access*” (Murray-rust, 2008), o movimento de acesso aberto. Exemplo disto, é a definição de *software* de código aberto, que utiliza frases que mapeiam quase totalmente a definição de dados abertos:

[...] podem ser utilizados, estudados e modificados sem restrições, que podem ser copiados e redistribuídos de forma modificada ou não modificada, quer sem restrições, ou com restrições apenas para garantir que outros destinatários também possam fazer estas coisas. [...] pode ser acompanhado por uma licença de *software* dizendo que o titular dos direitos de autor permite estes atos (uma licença [...] gratuita), ou ser liberado para o domínio público, de modo que estes direitos são automaticamente detidos (Murray-rust, 2008, p. 1, tradução própria).

Um importante aspeto dos dados abertos é que estes podem ser reutilizáveis, ou seja, a utilização dos dados ocorre normalmente sem autorização explícita, para estudos que estão previstos, ou não, pelo criador original. Estes então, podem ser utilizados para agregação, parâmetros em simulações e *mashups* onde podem ser combinados dados de diferentes fontes para dar origem a novas perspetivas (Murray-rust, 2008).

Uma das iniciativas mundiais que faz uso dos dados abertos é o European Data Portal (EDP). O EDP tem como finalidade recolher metadados do setor público que se encontram disponíveis em portais de dados abertos de vários países Europeus. Para além disto, é também objetivo do EDP promover o acesso e valorizar os dados abertos.

Os portais de onde são recolhidos os metadados podem ser nacionais, regionais, locais ou até mesmo de um domínio específico. No EDP existem secções para pesquisa de conjuntos de dados, fornecimento de dados, utilização de dados, formação e biblioteca.

Para além de recolher os metadados, o EDP realiza periodicamente uma avaliação do estado de maturidade dos dados lá depositados. Esta avaliação incide em quatro dimensões: política, portal de onde os dados são recolhidos, impacto e qualidade. Os dados dos diversos países são alinhados e os países são encaixados em quatro categorias de acordo com a sua maturidade: os líderes, os seguidores rápidos, os seguidores e os principiantes.

Assim como em outras iniciativas, há grande preocupação por parte do EDP quanto à qualidade dos dados. Embora o Portal adote uma avaliação abrangente, inclusive baseada nos Princípios FAIR (*Fairness, Accuracy, Interoperability and Reuse*), observa-se que alguns aspetos não são contemplados. Em virtude deste indicativo, buscar-se-á, a partir das recomendações de boas práticas apresentadas pelo *World Wide Web Consortium (W3C)* para dados na Web, analisar a qualidade dos conjuntos de dados disponibilizados pelo Portal.

A 31 de janeiro de 2017, o W3C lançou uma recomendação com 35 boas práticas¹ para os dados na Web (Data on the Web Best Practices – DWBP) (Lóscio et al., 2017). Este conjunto de boas práticas vem responder a vários desafios encontrados na publicação e à reutilização de dados.

¹ Embora a tradução literal para Best Practices seja melhores práticas, neste estudo será adotada a tradução oficial para o português do documento da W3C realizada por Lóscio et al. (2017).

1.2 Iniciativas

Quando se fala em abertura dos dados, há dois movimentos principais que fazem campanha por uma maior abertura de informações, documentos e conjuntos de dados que estão na posse de organismos públicos, sendo eles o movimento “Direito à informação” e o “Dados Abertos Governamentais”. O primeiro movimento reclama o direito público de acesso à informação e o movimento “Dados Abertos Governamentais”² utiliza argumentos sociais e económicos para incentivar a abertura de dados governamentais (Ubaldi, 2013).

Estes dois movimentos visam aumentar a transparência do governo de forma a que todos possam usufruir do valor social e económico gerado pela disponibilidade de informação (Ubaldi, 2013). Contudo, existem diferenças entre os dois, o movimento “Direito à informação” reclama o acesso à informação qualitativa e quantitativa armazenada em documentos, enquanto que o movimento “Dados Abertos Governamentais” reclama dados mantidos em bases de dados governamentais, preocupando-se com questões técnicas e jurídicas relacionadas ao acesso, à utilização e à reutilização dos dados (Ubaldi, 2013).

Cada vez mais governos de todo o mundo encontram-se a delinear ou a implementar estratégias de dados abertos com o objetivo de aumentar a transparência, participação e/ou eficiência governamental (Hoijsboom & Heeks, 2011).

Com a implementação destas estratégias, a publicação de dados governamentais num formato reutilizável “[...] pode reforçar o envolvimento dos cidadãos e produzir novas empresas inovadoras” (Hoijsboom & Heeks, 2011, p. 1, tradução própria).

Pode-se entender como dados governamentais todos os dados e informações originados ou encomendados por organismos públicos (Ubaldi, 2013). Contudo, é necessário perceber que existe uma diferença entre governo aberto e dados abertos governamentais. Um governo pode ser considerado um governo aberto, “[...] no sentido de ser transparente, mesmo que não abrace novas tecnologias, e um governo pode fornecer dados abertos e continuar a ser profundamente opaco e inexplicável” (Ubaldi, 2013, p. 14, tradução própria).

O termo “dados abertos governamentais” tem vindo a ganhar força sensivelmente a partir do ano de 2008, após a publicação de um conjunto de princípios de dados abertos governamentais nos Estados Unidos (EUA) (Ubaldi, 2013).

² Em inglês Open Government Data (OGD), que doravante será adotada como sigla para o termo para ambos os idiomas.

No primeiro dia do seu mandato como Presidente dos Estados Unidos da América, Barack Obama, em janeiro de 2009, anunciou que a sua administração iria iniciar uma estratégia de dados abertos, o que levaria a um nível de abertura do governo (Hoijsboom & Heeks, 2011).

Apesar das críticas, muitos países foram inspirados pela iniciativa adotada por Obama, o que levou, em dezembro de 2009, o Reino Unido a publicar um relatório onde declarava que o governo necessitava “[...] abrir radicalmente e promover a transparência [...]” (Hoijsboom & Heeks, 2011, p. 2, tradução própria). Em maio de 2010, seguiu-se o governo Australiano, que publicou uma Declaração de Governo Aberto (The Australian Government Information Management Office - AGIMO) em 2010 na qual a informação e o empenho dos cidadãos eram apoiados através de uma maior transparência governamental (Hoijsboom & Heeks, 2011). O governo Dinamarquês, em julho de 2010, lançou uma estratégia de inovação de dados abertos (*Offentlige Data I Spil*) (Hoijsboom & Heeks, 2011) e em várias regiões de Espanha foram desenvolvidas políticas de dados abertos. Com a Declaração de Visby, do Conselho Europeu, os estados membros da União Europeia deveriam “[...] procurar tornar os seus dados livremente acessíveis, em formatos legíveis por máquina e estimular a reutilização da informação do setor público” (Hoijsboom & Heeks, 2011, p. 2, tradução própria).

Destas iniciativas resultaram portais de dados abertos que permitem aos cidadãos e partes interessadas obter mais informações governamentais sobre determinado país (Attard et al., 2015).

A implementação de uma estratégia de dados abertos traz consigo vantagens e desvantagens, o que torna importante conhecer quais são as principais motivações/condutores da implementação e quais as barreiras enfrentadas quando esta iniciativa é instaurada.

Assim, os principais condutores à implementação de dados abertos são: as estratégias e experiências dos países pioneiros na implementação desta iniciativa, liderança política, iniciativas regionais, iniciativas dos cidadãos, iniciativas de mercado, tecnologias emergentes, legislação europeia, líderes do pensamento, possibilidade de monitorizar o governo e cortes orçamentais (Hoijsboom & Heeks, 2011).

As principais barreiras para a implementação de uma iniciativa de dados abertos são: uma cultura governamental fechada, legislação sobre privacidade e qualidade limitada dos dados.

Apesar das principais motivações para iniciativas de dados abertos mencionadas acima, a luta contra a corrupção, inicialmente, também constituía uma motivação para estas iniciativas, uma vez que esta prejudica a economia e a sociedade em todo o mundo. A transparência, que uma iniciativa de dados abertos proporciona, pode ser considerada “[...] como um desincentivo à corrupção” (Attard et al., 2015, p. 1, tradução própria).

A iniciativa de dados abertos governamentais tem a capacidade de aumentar a eficiência, eficácia e inovação do governo na entrega de serviços, bem como melhorar operações internas do setor público (Ubaldi, 2013). “[...] Os Dados Abertos Governamentais também podem ajudar a promover a colaboração entre e dentro de agências e departamentos públicos” (Ubaldi, 2013, p. 15, tradução própria).

Apesar de existir inúmeras iniciativas de dados abertos governamentais, também existe imensas barreiras, sendo a mais comum a barreira técnica que impede que estas iniciativas atinjam o seu potencial máximo.

Quanto ao ciclo de vida dos dados abertos governamentais, este pode ser entendido como tendo três etapas: a etapa de pré-processamento, a etapa de exploração e a etapa de manutenção, sendo respetivamente cada etapa responsável por preparar os dados a publicar, utilizar os dados publicados e manter os dados publicados de forma que estes sejam sustentáveis (Attard et al., 2015).

No sentido de atingir um maior crescimento económico, as entidades governamentais devem ter em atenção as necessidades dos utilizadores finais, sendo que os dados fornecidos devem ser abertos num formato legível por máquina e não proprietário. O W3C recomenda a utilização de normas e ferramentas abertas estabelecidas, como o XML (*Extensible Markup Language*) e RDF (*Resource Description Framework*), como formato da publicação dos dados (Attard et al., 2015), entre outras. Tópicos a serem explorados a partir do Capítulo 1 desta dissertação.

1.3 Problema/Questões

O EDP realiza periodicamente uma avaliação do estado de maturidade dos dados, avaliação esta que incide em quatro dimensões: política, portal, impacto e qualidade. Apesar

de ampla, esta avaliação não é suficientemente profunda em relação a aspetos como a qualidade dos dados.

O EDP adota uma avaliação da qualidade dos dados baseada nos Princípios FAIR, uma vez que, para os dados estarem de acordo com os princípios FAIR, estes têm de ser: localizáveis, acessíveis, interoperáveis e reutilizáveis. Assim, o EDP analisa: a utilização de palavra-chave, as categorias, se permite pesquisa temporal, pesquisa espacial, a acessibilidade da URL de acesso, a URL de transferência, a acessibilidade da URL de transferência, o formato, o tipo de suporte, se tem distribuições não proprietárias, se tem distribuições legíveis por máquina, se está em conformidade com o sistema DCAT-AP³, se tem licença, quais as restrições, tipo de vocabulário, os pontos de contacto, o publicador, o tamanho do ficheiro, a data de emissão e de alteração.

Em vista da proposta deste estudo em analisar conjuntos de dados do EDP à luz das Boas Práticas para Dados na Web (Lóscio et al., 2017), são legítimos, portanto, as seguintes questões de pesquisa: Em que medida os conjuntos de dados presentes no EDP cumprem as boas práticas para dados na Web recomendadas pelo W3C? Quais as boas práticas atendidas com maior e menor frequência pelos conjuntos de dados destas categorias?

De modo a delimitar o âmbito de pesquisa, centra-se a análise em 3 categorias dentre as 13 existentes no EDP, sendo elas as seguintes: de Governo e Setor Público, Transportes e Educação, Cultura e Desporto, presentes no EDP, pelo que as questões a que pretende responder são as seguintes:

- a) Em que medida os conjuntos de dados nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto, presentes no European Data Portal cumprem as boas práticas para dados na Web recomendadas pelo W3C?; e
- b) Quais as boas práticas atendidas com maior e menor frequência pelos conjuntos de dados destas categorias?

As boas práticas para os dados na Web foram lançadas pelo W3C há relativamente pouco tempo e nunca foram utilizadas antes para aferir a qualidade dos dados do EDP. Quando pesquisado por artigos que citavam as DWBP, pesquisa de busca por “Data on the

³ DCAT Application Profile para portais de dados na Europa.

Web Best Practices, W3C”, “DWBP”, “recommendations W3C” (análise realizada no dia 8/Dez/2020) apenas foram encontrados 116 resultados no “*Google Scholar*” e 4 resultados no “*Scopus*”. Ao todo, estes 120 resultados encontrados, apenas dois se referiam especificamente à implementação das boas práticas para a Web, enquanto os outros apenas citam as DWBP. Os artigos analisados que implementam as DWBP são: “*Implementation of the Best Practices for Data on the Web in Brazil and Costa Rica*” de (Lóscio et al., 2019) e “*Um Survey sobre Soluções para Publicação de Dados na Web sob a Perspetiva das Boas Práticas do W3C*” de (Oliveira et al., 2017).

Estes artigos apesar de se referirem especificamente às boas práticas para a Web, não as utilizam para fazer a análise de qualidade de dados, mas sim para exemplificar os passos seguidos pelo Brasil e Costa Rica para implementar as DWBP de forma a melhorar a qualidade do processo de publicação de dados no centro de estudos sobre o desenvolvimento da sociedade de informação (Cetic.br) e no departamento judiciário da Costa Rica. No caso do “*Um Survey sobre Soluções para Publicação de Dados na Web sob a Perspetiva das Boas Práticas do W3C*” é pesquisado sobre as soluções mais utilizadas atualmente para a publicação de dados na Web e comparar se estas estão alinhadas com as boas práticas recomendadas pelo W3C.

Na revisão da literatura, ficou claro que existe uma lacuna na literatura em relação a estudos sobre as boas práticas para a Web (W3C), até mesmo pelo reduzido número de estudos onde a qualidade dos conjuntos de dados seja avaliada segundo as boas práticas.

1.4 Objetivo Geral e Específicos

A realização desta dissertação tem como objetivo principal determinar em que medida os conjuntos de dados das categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto publicados no EDP atendem às boas práticas de publicação de dados na Web recomendadas pelo W3C.

1.4.1 Objetivos Específicos

Para consecução do objetivo desta investigação, são propostos os seguintes objetivos específicos:

- a) estudar as dimensões analisadas pelo EDP;

- b) comparar os indicadores das dimensões do EDP com os indicadores DWBP;
- c) identificar os conjuntos de dados a analisar das 3 categorias selecionadas;
- d) analisar e classificar os conjuntos de dados com base nas DWBP; e
- e) determinar quais as boas práticas atendidas com maior e menor frequência pelos conjuntos de dados.

Esta análise torna-se relevante, uma vez que vai aprofundar a análise de qualidade dos dados do *European Data Portal*, e irá cobrir uma lacuna existente na literatura no que se refere à análise da qualidade dos dados presentes no portal como um todo, neste caso o EDP, à luz das DWBP.

1.5 Estrutura da Dissertação

Esta dissertação está estruturada em cinco capítulos, além da Introdução. Na Introdução são abordadas as principais iniciativas de dados abertos, explica-se o problema/questões e motivação para a realização deste trabalho, refere-se o objetivo geral da dissertação e os objetivos específicos.

No Capítulo 2, ocorre a fundamentação teórica sobre Dados Abertos, Melhores práticas de Dados Abertos, Dados Abertos Governamentais, EDP, princípios FAIR e DWBP.

No terceiro Capítulo, refere-se ao enquadramento metodológico. Segue-se uma fundamentação teórica sobre o tipo de estudo e o tipo de abordagem adotada e os procedimentos metodológicos utilizados.

O quarto Capítulo diz respeito à análise e à discussão dos resultados. Efetua-se uma reflexão crítica sobre os resultados e sobre estes à luz da informação constante noutros estudos, em particular às análises efetuadas pelo EDP.

Por último, o Capítulo 5 diz respeito às conclusões do estudo, onde se responde às questões colocadas no Primeiro capítulo e considerações sobre os resultados decorrentes das análises realizadas.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Dados Abertos

A tecnologia tem um papel fundamental nas mudanças que ocorrem na sociedade e o surgimento da Internet é a maior mudança tecnológica que já existiu. Com a Internet, é possível ter acesso a enormes quantidades de informação de forma imediata, o que faz com que empresas possam desenvolver novos produtos e serviços.

Nos últimos anos, vários desafios têm surgido frente ao grande volume de informação, sobretudo ao de dados. No contexto da Web, tem sido abordado conceitos, características, regulamentações, formas de divulgação e uso dos dados. Como dados abertos entende-se toda a informação que pode ser livremente utilizada, modificada e compartilhada por qualquer pessoa para qualquer fim (Carrara et al., 2015), que, quando publicado, é acompanhado de uma licença aberta e serem legíveis por máquina.

Já em 2015 era previsto que entre os anos de 2016 e 2020, o volume de dados abertos crescer por volta de 36,9%, atingindo um valor de 75,7 mil milhões de euros no ano de 2020 (Carrara et al., 2015).

Para avaliar o nível de maturidade de uma organização, ou de um Estado em relação aos dados abertos, têm surgido vários modelos que abordam aspetos como a qualidade dos dados. Para avaliar a qualidade dos dados é possível encontrar, por exemplo, o modelo de cinco estrelas dos dados ligados (*Linked Data*), proposto por Tim Berners Lee. Este modelo avalia o grau de estruturação dos dados em direção aos dados abertos e permite classificar os dados até cinco estrelas. Para conseguir a classificação de uma estrela, os dados devem estar disponíveis na Internet com uma licença aberta. Para a classificação de duas estrelas, os dados devem estar estruturados e possuir formatos legíveis por máquina. Para a classificação de três estrelas, os dados devem ter as mesmas características dos níveis anteriores e estar também em formato aberto. Para atingir a classificação de quatro estrelas, os dados devem ter as características anteriores e ainda usar padrões abertos do W3C para identificação. Para alcançar o máximo de cinco estrelas, os dados devem ter tudo o que foi referido anteriormente, mas também referir dados de terceiros de forma a garantir maior contextualidade (Dados.gov.br., 2021).

A utilização de dados abertos contribui para que as organizações consigam entender melhor os seus clientes, bem como as suas escolhas, ajudando-os a tomar decisões mais inteligentes e eficazes. No que diz respeito ao processo operacional, está demonstrado que as organizações podem aumentar a sua performance usando ferramentas digitais para tomada de decisão com base nos dados abertos (Carrara et al., 2015).

Mas, nem só as organizações e grandes empresas estão a perceber a relevância do papel dos dados abertos, as cidades e os seus cidadãos também. Os dados abertos podem ser bastante úteis, por exemplo, na utilização de transportes, no sentido de ajudar a identificar locais de acidente, mas também a ter uma resposta mais eficiente em emergências, como por exemplo, “calcular a probabilidade de ocorrência de um incêndio numa área específica” (Carrara et al., 2020, p. 5, tradução própria).

Com a abertura dos dados na Web surgiram tanto novas oportunidades como novos desafios para editores e consumidores de dados, como por exemplo a “[...] forma de representar, descrever e disponibilizar dados (pelos editores) de uma forma que seja facilmente encontrados, compreendidos, utilizados e reutilizados (pelos consumidores)” (Zeng et al., 2017, p. 1, tradução própria).

Os dados abertos são vistos como um “[...] recurso essencial para o crescimento económico, a criação de emprego e o progresso da sociedade [...]” (Berends et al., 2017, p. 7, tradução própria). Os dados abertos trazem inúmeros benefícios, tais como: fornecem uma visão que ajuda na tomada de decisão, seja esta de forma visualizada ou por referência e ajuda a perceber a importância da reutilização de dados. O setor que mais se beneficia com a abertura de dados é o setor público, o que leva a concluir que o setor público é o primeiro reutilizador dos seus próprios dados (Berends et al., 2017) .

Nesse contexto, apresenta-se a Public Sector Information (PSI), como sendo os dados que podem ser reutilizáveis, mas é preciso ter em consideração em que condições tais reutilizações podem acontecer (Huyer & Cecconi, 2018).

O processo de publicação e reutilização de PSI passa primeiramente por publicar a PSI, seguido de encontrar e compreender a PSI e finalmente reutilizar a PSI. Quando a PSI é publicada sob uma licença aberta, esta passa a ser considerada como Dados Abertos Governamentais (Huyer & Cecconi, 2018). Lóscio et al. (2017) e Torino et al. (2019) acrescentam que além de estabelecer uma licença, deve-se indicar o tipo de licença e a versão da licença.

De forma a aumentar mais ainda os benefícios que os dados abertos trazem para a sociedade e a economia, os esforços devem estar centralizados nos conjuntos de dados que possuem maior potencial.

[...] High-value datasets significa documentos cuja reutilização está associada a importantes benefícios para a sociedade, o ambiente e a economia, em particular, devido à sua adequação para a criação de valor acrescentado para serviços, aplicações e novos empregos de alta qualidade e decentes, e do número de potenciais beneficiários dos serviços e aplicações de valor acrescentado baseados naqueles datasets [...] (Huyer & Blank, 2015, p. 3, tradução própria).

Para oficializar a publicação de dados abertos enfrenta-se também muitas barreiras, pelo que é essencial tornar este processo “[...] parte integrante do processo de criação de dados, em vez de uma atividade separada ou adicional aos processos e rotinas operacionais diários” (Berends et al., 2017, p. 11, tradução própria).

Uma barreira importante é “[...] a falta de competências entre vários departamentos governamentais e níveis inferiores de governo para lidar com dados abertos” (Berends et al., 2017, p. 13, tradução própria). Para além desta, destaca-se a barreira legal, nomeadamente os “desafios relacionados com a legislação de dados abertos, políticas de dados abertos, diretivas e licenças governamentais” (Berends et al., 2017, p. 16, tradução própria). Soma-se ainda, a barreira financeira e técnica, principalmente no que diz respeito à qualidade dos metadados, foco deste estudo.

2.2 Dados Abertos Governamentais

As iniciativas de dados abertos governamentais têm como finalidade aumentar a transparência, participação, colaboração, inovação, melhorar o valor económico e criar valor público (Kawashita et al., 2020). “Os dados governamentais são vistos como um recurso valioso que oferece benefícios potenciais para partes interessadas se estiver disponível abertamente” (Kawashita et al., 2020, p. 2, tradução própria). No entanto, estes dados só têm realmente valor quando são reutilizados.

Em dezembro de 2007 foram apresentados oito princípios de dados abertos governamentais (OGD), numa reunião de trabalho na Califórnia que reuniu cerca de 30 defensores dos dados governamentais abertos (Ubaldi, 2013). Os oito princípios são:

completos, primários, atuais, acessíveis, processáveis por máquina, não discriminatórios, não proprietários e licença livre (Ubaldi, 2013).

Segundo Lee & Kwak (2011), as iniciativas de dados abertos governamentais são incrementais, pelo que sugerem um modelo de quatro fases para que os dados governamentais se tornem totalmente abertos: na etapa zero, é necessário obter uma visão dos dados governamentais que já existem; na etapa um, deve ser aumentada a transparência dos mesmos; na segunda etapa, é importante melhorar a participação aberta; na terceira, deve-se aumentar a colaboração aberta; e na última e quarta, deve-se realizar um envolvimento omnipresente (Ubaldi, 2013).

Várias iniciativas para a reutilização de dados abertos governamentais estão a surgir nos países da Organização para a Cooperação e Desenvolvimento Económico (OCDE). Por exemplo, as *Apps for Democracy*, lançadas pelo governo dos USA, “[...] aparentemente levou a um retorno estimado de 4.000% em investimento num período de 30 dias” (Ubaldi, 2013, p. 26, tradução própria). Outra iniciativa semelhante, na Noruega, o *Nettskap 2.0*, resultou no desenvolvimento de 135 aplicações informáticas (Ubaldi, 2013).

A Carta Internacional de Dados Abertos foi criada em 2016 para verificar como as avaliações dos aspetos da publicação e utilização do OGD estão a ser cumpridas. Os princípios desta carta proporcionam uma abordagem consistente para avaliar o grau de abertura dos dados e devem ser entendidos como compromissos, que os governos e as instituições devem assumir com o objetivo de “[...] melhorar a qualidade dos dados publicados e as condições de partilha e reutilização de dados abertos” (Kawashita et al., 2020, p. 9, tradução própria).

A avaliação dos dados abertos baseia-se na recolha, análise e fornecimento de informações relativas às iniciativas de dados abertos. “Os instrumentos de referência do OGD podem ser utilizados para avaliar e classificar países, organizações e projetos com base na forma como publicam e utilizam o OGD de diferentes formas” ((Kawashita et al., 2020, p. 4, tradução própria). O nível de análise de avaliação do OGD varia de instrumento para instrumento.

A maioria das avaliações vigentes baseiam-se a nível do país, enquanto as avaliações de implementação podem ser por conjuntos de dados individuais, portais, instituições individuais, iniciativas do OGD e países inteiros (Kawashita et al., 2020).

Uma análise com seis instrumentos de referência⁴ mostra que estes foram comparados segundo o nível de análise, cobertura geográfica, frequência, abordagem metodológica e avaliação, recolha de dados/fonte de dados, conceitos mencionados e cobertura dos princípios do *International Open Data Chapter*. Esses instrumentos foram criados para responder a diferentes objetivos, com diferentes graus de especificidade, alcance e foco. Os seis princípios presentes na *International Open Data Chapter* são abrangidos pelos seis instrumentos de referência, diferindo apenas no foco que cada um oferece ((Kawashita et al., 2020).

Vale salientar que “A maioria dos instrumentos de referência baseiam-se em inquéritos para recolher dados e validar resultados com a ajuda de peritos, funcionários governamentais e da comunidade” ((Kawashita et al., 2020, p. 7, tradução própria).

2.3 Melhores Práticas de Dados Abertos Governamentais

Em relação às iniciativas de dados abertos, existem países que se destacam em relação às melhores práticas de dados abertos, sendo considerados até como tendo as melhores *performers* neste campo. Na Europa, países como Irlanda, Espanha e França são exemplos de como alcançar uma gestão bem-sucedida de dados abertos. A análise realizada em relação a estes países está de acordo com as dimensões de avaliação dos Dados Abertos Pan-Europeus de 2018: política, portal, impacto e qualidade de dados abertos (Radu, 2020).

Quando analisada a política de dados abertos nestes três países, é possível identificar fatores comuns de sucesso, como por exemplo: ter uma visão estratégica em torno dos dados abertos e possuir uma estrutura de governação de dados aberta e inclusiva.

No que diz respeito à compreensão sobre âmbito e missão do portal nacional de dados abertos, estes três países são avaliados como casos bem sucedidos. Para além desta compreensão, Irlanda, Espanha e França possuem uma imagem do seu portal de dados abertos como algo que vai para além de um repositório de dados. Os portais de dados destes países estão a evoluir para “espaços virtuais de desenvolvimento, troca e interação de conhecimentos sobre Dados Abertos” (Radu, 2020, p. 11, tradução própria). Assim, nos portais de dados destes três países foram encontrados dois pontos em comum:

⁴ European Open Data Maturity Assessment (EODMA), Open-Useful-Re-Usable Government Data Index (OURdata), Open Data Barometer (ODB), Global Open Data Index (GODI), Open Data Inventory (ODIN) e Open Data Monitor (ODM).

[...] características avançadas que promovem o acesso aos dados e permitem a interação entre utilizadores; acesso a dados, ferramentas e conteúdos editoriais para atender um vasto público, composto por editores, reutilizadores de todos os setores e do público interessado em geral [...] (Radu, 2020, p. 11, tradução própria).

Relativamente à qualidade dos dados, a Irlanda, Espanha e França possuem um papel diferenciador nesta questão. Estes países mostram uma abordagem sistemática para “[...] assegurar a capacidade de descoberta e atualização regular dos dados abertos provenientes de dados locais, regionais e fontes nacionais; e controlar a qualidade e promover uma melhor abertura na publicação de dados no país” (Radu, 2020, p. 19, tradução própria).

Os países Chipre, França e Irlanda, relativamente à política, portal, impacto e qualidade dos dados, mostram ter as melhores práticas, o que pode ajudar outros países a alcançarem o mesmo patamar, estando sempre dependentes da situação em que cada país se encontra (Radu, 2020).

No que diz respeito aos portais de dados abertos, estes estão cada vez mais a permitir encontrar conjuntos de dados, a possibilitar a “[...] interação entre editores e reutilizadores de dados através de fóruns e o *feedback* dos dados e sistemas de classificação” (Knippenberg, 2015, p. 31, tradução própria) . Os três países consideram que o portal de dados abertos deve incluir os dados públicos e privados e tipos diferentes de informação de forma a melhorar a experiência dos utilizadores.

Os três países desempenham mais esforço e dedicação no que toca à qualidade dos dados e metadados. Para fornecer apoio, existem manuais, páginas de FAQ e diretrizes, que propiciam o *feedback* dos editores e reutilizadores.

2.4 European Data Portal

Como portal de dados é possível entender como sendo interfaces na Web, planeadas para facilitar a localização de informações que podem ser reutilizáveis (European Commission, 2021). Os portais de dados contêm metadados dos conjuntos de dados publicados para reutilização. Para além disto, oferecem funcionalidades de pesquisa específica, facilitam a descoberta de conjuntos de dados de maior interesse e disponibilizam frequentemente

*Application Programming Interfaces (APIs)*⁵ para acesso direto e automatizado aos dados (European Commission, 2021).

Os portais de dados representam um papel fundamental nas iniciativas de dados abertos, sendo utilizados principalmente pelas administrações públicas a nível europeu, nacional e local dos Estados-Membros (European Commission, 2021).

Os principais objetivos dos portais de dados abertos é aumentar a transparência, a participação e eficiência dos governos. De forma a garantir a continuidade da publicação dos dados, é importante garantir o futuro dos portais de dados abertos. As dez formas para os portais de dados abertos evoluírem atingindo a sustentabilidade e o valor acrescentado são as seguintes: “[...] ser detetável, ser mensurável, promover a utilização, organizar para uso, ser acessível, promover padrões, publicar metadados, proporcionar dados de ligação, co-localizar documentação e disponibilizar ferramentas de co-localização” (Simperl & Walker, 2017, p. 7, tradução própria).

De forma a avaliar se um determinado portal de dados abertos cumpre com as normas de publicação de dados sustentáveis e reutilizáveis, utiliza-se certificados de dados.

Para avaliar um portal de dados abertos, utilizam-se pelo menos dois tipos de métricas: a utilização (para editores) e a qualidade (para os utilizadores). A métrica de *downloads* é a métrica mais usada, mas, no entanto, não explica para que propósito os dados estão a ser utilizados. Para avaliar a qualidade de um portal não existe uma definição padrão, porém tende-se a distinguir entre critérios absolutos (exaustividade, correção e atualidade) e critérios relativos, como por exemplo a aptidão de utilização (Simperl & Walker, 2017).

Os dados devem ser acompanhados de metadados, descrições de relevância, usabilidade e qualidade, comentários, casos de uso e exemplos. Os metadados acrescentam valor aos dados, mas também têm um grande potencial para melhorar a reutilização (Simperl & Walker, 2017).

Dentre os portais de dados abertos governamentais existentes na Europa, pode-se citar: Paris Data (opendata.paris.fr), www.data.gouv.fr, Open Data della Regione Piemonte (www.dati.piemonte.it), www.dati.gov.it, Open data van de Overheid

⁵ Interfaces de Programação de Aplicativos

(www.data.overheid.nl), o foco deste estudo passa por analisar o European Data Portal (EDP) (<https://data.europa.eu/en>) (European Commission, 2021).

O European Data Portal (EDP) foi disponibilizado em 2016. O EDP recolhe metadados disponíveis em dados públicos e portais geoespaciais nos países europeus, o que inclui países membros da UE, países da EFTA e países envolvidos na política de vizinhança da UE. Os conjuntos de dados incluem, por exemplo, registos de terrenos, mapas de estado, e a localização das estações dos correios. O acesso ao Portal é possível através de API legíveis por máquina e de website legível por humanos National Spatial Data Infrastructure., 2021). Para além disso, o Portal também fornece treze categorias de dados definidas de acordo com os domínios Eurovoc. Este thesaurus permite aos utilizadores realizar pesquisas multilingues por categorias de dados (Data.Europa.Eu, 2021, Lóscio et al., 2017).

As categorias de conjuntos de dados existentes no EDP são Agricultura, Pesca, Silvicultura e Alimentação, Ambiente, Ciência e Tecnologia, Economia e Finanças, Educação, Cultura e Desporto, Energia, Governo e Setor Público, Justiça, Sistema Judiciário e Segurança Pública, População e Sociedade, Questões Internacionais, Regiões e cidades, Saúde e Transportes.

O EDP visa também promover o acesso e o valor dos dados abertos. Tal como acontece com outras iniciativas, existe uma grande preocupação por parte da EDP no que diz respeito à qualidade dos dados. Embora o Portal adote uma avaliação abrangente baseada nos Princípios FAIR, alguns aspectos não são contemplados.

Este Portal realiza periodicamente uma avaliação do estado de maturidade dos dados, no sentido de perceber o estado de desenvolvimento alcançado na área dos dados abertos na Europa.

O âmbito da avaliação da maturidade dos dados aborda quatro dimensões:

- a) política de dados abertos, que diz respeito à presença de políticas e estratégias específicas, no sentido de promover dados abertos a nível nacional;
- b) o portal de dados abertos, ou seja, as funções que o portal de dados apresenta, de forma que utilizadores mais experientes e também os menos experientes, consigam aceder aos dados através do portal e também, características que possam melhorar a interação entre editores e reutilizadores de dados;

- c) impacto de dados abertos, ou seja, verifica-se as atividades realizadas no sentido de monitorar e medir a reutilização e o impacto de dados abertos. Esta dimensão concentra-se em quatro áreas de impacto dos dados: política, social, ambiental e económica; e
- d) qualidade dos dados abertos, diz respeito às medidas adotadas pelos gestores do portal no sentido de assegurar a recolha metódica de metadados, através de fontes de todo o país. Também se preocupa com a interoperabilidade dos dados ao adotar um formato comum de metadados (*Data Catalog Vocabulary – Application Profile*) e o uso do *Resource Description Framework (RDF)*, bem como a qualidade dos dados publicados (European Data Portal, 2020).

Relativamente à política de dados abertos, todos os 28 países presentes na European Union (EU28) no ano 2019 adotaram uma política relativa aos dados abertos. Em vários países a política de dados abertos centra-se na agenda digital nacional, estratégia digital, estratégia nacional de dados, ou até mesmo no contexto de governo aberto. No que diz respeito, à adoção de uma estratégia de dados abertos, 23 países do EU28 adotaram uma estratégia. No entanto, estas estratégias nem todas são exclusivas aos dados abertos, muitas delas são incorporadas no contexto de Governo Aberto (European Data Portal, 2020). A dimensão política nos países pertencentes a EU28 apresentam um nível de maturidade relativamente alto, situando-se nos 74%. O indicador com mais maturidade dentro da dimensão política é a governança dos dados abertos (79%) (European Data Portal, 2020).

Quanto ao portal de dados abertos, o EDP, foca-se na maturidade que os portais de dados abertos apresentam. Para avaliar esta dimensão, utilizam-se as seguintes métricas: características do portal, utilização do portal, disposição dos dados e sustentabilidade do portal.

Para facilitar a relação entre editores e reutilizadores de dados, os portais de dados em geral devem apresentar as seguintes características: possuir um canal de *feedback*, sendo que 27 dos portais da EU28 apresentam esta característica, ter um fórum de discussão (sendo que 16 portais europeus têm esta característica) e permitir receber notificações quando novos conjuntos de dados são disponibilizados (19 dos EU28) (European Data Portal, 2020).

Apenas 12 dos 28 portais de dados abertos europeus realizaram um inquérito aos seus utilizadores para perceber a sua satisfação em relação ao mesmo. Para melhorar os portais, 23 dos EU28 países têm um processo em vigor que passa pelo portal ser revisto e melhorado

regularmente, sendo que 39% são revistos trimestralmente, 26% são revistos bianualmente, 17% são revistos anualmente e 17% ainda menos frequentemente (European Data Portal, 2020).

Assim, o nível de maturidade da dimensão portal é de 67%, sendo que o indicador com mais maturidade nesta dimensão é a utilização do portal e o indicador com menos é a sustentabilidade do portal (European Data Portal, 2020).

“O impacto dos dados abertos analisa as abordagens e metodologias existentes desenvolvidas a nível de país e organismo público para controlar e medir a reutilização e o impacto dos dados abertos” (European Data Portal, 2020, p. 37, tradução própria). Para avaliar esta dimensão são utilizadas as seguintes métricas: sensibilização estratégica, impacto político, impacto social, impacto ambiental e impacto económico.

O nível de maturidade desta dimensão é de 57%, sendo o nível de maturidade mais baixo das quatro dimensões de dados abertos. O indicador que apresenta maior maturidade dentro da dimensão é o impacto ambiental (69%), em oposição com o impacto económico dos dados que apresenta menor maturidade (28%) (European Data Portal, 2020).

A qualidade dos dados abertos, como já foi referido foca-se nas medidas adotadas pelos gestores do portal, de forma a garantir a recolha sistemática de dados em todo o país, a disponibilidade dos metadados disponíveis, dados reais e garantir a conformidade com o DCAT-AP (*Data Catalog Vocabulary- Application Profile*). De forma a avaliar esta dimensão, são utilizados os seguintes indicadores: disponibilidade, monitorização e medidas, conformidade DCAT-AP e qualidade de implantação e dados ligados.

Possuir uma licença livre sobre os dados é muito importante, uma vez que sem licença os utilizadores ficam sem saber se os dados podem ser usados e partilhados. Assim, apenas 54% dos Estados Membros possuem todos os seus conjuntos de dados com licença. Para ajudar os editores de dados a escolher uma licença para os seus dados, foi recomendado por 21 países dos EU28 a utilizarem a Licença *Creative Commons (CC)* (European Data Portal, 2020).

Referindo-se ao indicador de qualidade de implantação e dados ligados, este indicador avalia até que ponto os países utilizam o modelo *5-Star Open Data*, de forma a avaliar a qualidade dos seus dados. Como referido anteriormente, este modelo permite aos países classificar a qualidade dos seus dados até 5 “estrelas”.

Para obter o máximo de 5 estrelas, os dados devem estar disponíveis online sob uma licença aberta, estar sob a forma de dados estruturados, ser legível por máquina, utilizar URIs como seus identificadores e incluir ligações a outras fontes de dados (European Data Portal, 2020, p. 57, tradução própria).

De acordo com este modelo, a maioria dos dados disponíveis nos portais de dados nacionais estão situados nas primeiras três estrelas do modelo, sendo que as estrelas quatro e cinco ainda não são comuns na Europa (European Data Portal, 2020).

Assim, o nível de maturidade desta dimensão situa-se nos 65%. O indicador com mais maturidade dentro desta dimensão é a monitorização e medidas (76%), ao contrário do indicador com menos maturidade que é o indicador de qualidade de implantação e dados ligados (53%) (European Data Portal, 2020).

No que diz respeito à maturidade dos dados abertos nos países da Associação Europeia de Comércio Livre (EFTA), ou seja, Islândia, Liechtenstein, Noruega e Suíça, a avaliação é feita sob as mesmas dimensões que nos Estados Membros da EU28, querendo isto dizer que a maturidade dos dados abertos destes países também é medida segundo as dimensões da política, portal, impacto e qualidade, sendo que as dimensões de política e portal de dados abertos apresentam uma maior maturidade, enquanto as dimensões do impacto e qualidade apresentam menor maturidade.

2.5 Princípios FAIR

Após estudar quais as dimensões analisadas pelo EDP para avaliar a qualidade dos conjuntos de dados, observou-se que estas estão relacionadas com os princípios FAIR. Por essa razão, esses princípios serão abordados neste tópico.

Os princípios FAIR, acrônimo adotado para *Fairness, Accuracy, Interoperability and Reuse*, foram criados em 2014, com o objetivo de “[...] fornecer um conjunto de princípios orientadores e práticas mínimas para os dados de investigação” (Boeckhout et al., 2018, p. 1, tradução própria). A adoção do FAIR possibilita que os dados sejam legíveis por máquinas e por seres humanos. (Dunning et al., 1970).

“Uma mensagem chave dos princípios FAIR é que os metadados e os padrões de metadados devem ser articulados e tornados publicamente disponíveis na maior medida possível” (Boeckhout et al., 2018, p. 2, tradução própria).

Outro aspeto dos princípios FAIR é que estes exigem que os dados sejam bem explícitos e definidos, bem como todos os termos e condições sob os quais estes são partilhados ou tornados acessíveis estejam prontamente disponíveis.

“Os princípios FAIR pretendem principalmente proporcionar condições para facilitar a partilha de dados, incluindo as condições de obtenção e concessão de acesso aos dados, privacidade, publicação e utilização” (Boeckhout et al., 2018, p. 3, tradução própria).

Para que os dados estejam alinhados aos princípios FAIR, eles devem ser localizáveis, acessíveis, interoperáveis e reutilizáveis. Ou seja, para que um objeto de dados seja localizável, este deve ser único, estar associado a um identificador persistente e acionável. Para que sejam acessíveis, os dados devem ser sempre recuperáveis por máquinas e humanos. Os objetos de dados são interoperáveis apenas se os metadados forem acionáveis por máquina; se os formatos dos metadados utilizarem vocabulários e/ou ontologias partilhados; e os metadados que se encontram dentro do objeto de dados forem sintaticamente analisáveis e semanticamente acessíveis por máquina. Para que os objetos de dados sejam reutilizáveis, os metadados devem ser suficientemente bem descritos, de maneira a que sejam facilmente vinculados ou integrados a outras fontes de dados e, devem ainda, referir a sua fonte e proveniência para futura citação (FORCE11, 2015).

Assim, conclui-se que o EDP ao verificar a pesquisabilidade, a acessibilidade, a interoperabilidade e a reusabilidade dos objetos de dados, vai ao encontro dos princípios FAIR. Logo, o EDP segue as dimensões do FAIR e inclui a dimensão contextualidade.

No próximo tópico serão apresentadas as boas práticas para os dados na Web (DWBP), indicadores a serem adotados para análise do EDP.

2.6 Boas Práticas para os Dados na Web (DWBP)

A 31 de janeiro de 2017, foi emitido pelo W3C um conjunto de boas práticas para a Web, que quando aplicadas trazem benefícios na forma como os conjuntos de dados estão disponíveis na Web.

A adoção das DWBP, recomendadas pelo W3C, melhora a compreensão dos dados, uma vez que, os humanos poderão ter uma melhor perceção sobre a estrutura, o significado, os metadados e a natureza dos conjuntos de dados. Melhora, ainda, a processabilidade, pois as máquinas serão capazes de processar e manipular automaticamente os dados do conjunto de

dados e descobrir automaticamente um conjunto de dados ou dados dentro de um conjunto de dados. Aumenta a possibilidade de reutilização do conjunto de dados por diferentes grupos de consumidores, assim como a confiança dos mesmos no conjunto de dados. Além de permitir criar *links* entre recursos de dados, os humanos e as máquinas podem aceder a dados atualizados em vários formatos, facilitando o consenso entre editores e consumidores de dados. Essas características estão associadas ao ciclo de vida dos dados na Web.

O ciclo de vida de dados na Web passa por sete estados, sendo eles os seguintes: planeamento, criação, publicação, acesso, consumo, *feedback* e aprimoramento (Burle, 2021). O ciclo de vida inicia-se no planeamento, onde se escolhe quais os dados que serão publicados. Segue-se a criação, onde se extrai os dados das fontes originais, procede-se à sua transformação e posteriormente a criação de metadados. Na fase de publicação, os dados e os metadados são disponibilizados através de catálogos, APIs, etc. A fase do acesso ocorre quando os consumidores têm acesso aos dados. O consumo inicia-se quando os dados são reutilizados, seja para a criação de novos dados, ou seja para visualizações, análises e aplicações. O estado de *feedback* permite aos consumidores de dados darem *feedback* sobre os dados e metadados. Por último, na fase de aprimoramento procede-se à manutenção dos dados, com correções e atualizações (Burle, 2021).

Atua neste contexto o grupo de trabalho das boas práticas para dados na Web, que tem como missão desenvolver o ecossistema de dados abertos, fornecer orientação aos publicadores e aumentar a confiança dos consumidores sobre os dados publicados (Burle, 2021).

“As Boas práticas abrangem diferentes aspetos relacionados à publicação e ao consumo de dados, como formatos de dados, acesso aos dados, identificadores de dados e metadados” (Lóscio et al., 2017).

Enquanto o esquema de cinco estrelas (Berners-Lee, 2006), por exemplo, espera apresentar o grau de abertura do formato dos dados, a recomendação do W3C, DWBP, apresenta práticas mais amplas, centradas nas capacidades técnicas da arquitetura Web (Penteado et al., 2019). Apresenta, ainda, como os dados:

[...] são entregues, codificados, mantidos e como se dá o relacionamento entre os produtores e os consumidores desses dados, estendendo os princípios dos dados abertos governamentais. Esta recomendação tem como prática geral o reaproveitamento de padrões já estabelecidos na Web e para a recuperação das informações (Penteado et al., 2019, p. 5, tradução própria).

A seguir serão detalhadas as práticas recomendadas pela W3C (Lóscio et al., 2017):

Melhor prática 1: fornecer metadados, passa por fornecer metadados para utilizadores humanos e aplicações de computador, de forma que os humanos possam entender os metadados e as aplicações de computador. “Os metadados fornecem informações adicionais que ajudam os consumidores de dados a entender melhor o significado dos dados, a sua estrutura e esclarecer outras questões, como direitos e termos de licença, a organização que gerou os dados, a qualidade dos dados, os métodos de acesso aos dados e o cronograma de atualização dos conjuntos de dados”.

Para fornecer metadados legíveis por humanos, é necessário fornecê-los como parte de uma página web HTML ou como arquivo de texto separado. Para os mesmos estarem legíveis por máquina, os metadados devem ser fornecidos em formato de serialização, como Turtle e JSON, ou incorporados na página HTML usando HTML-RDFA⁶ ou JSON-LD⁷.

Melhor prática 2: fornecer metadados descritivos, passa por fornecer metadados que descrevem as características gerais dos conjuntos de dados e suas distribuições, facilitando a descoberta dos mesmos na web e sua natureza. Os metadados descritivos de um conjunto de dados, podem incluir as seguintes características: “O título e uma descrição do conjunto de dados; As palavras-chave descrevendo o conjunto de dados; A data de publicação do conjunto de dados; A entidade responsável por disponibilizar o conjunto de dados; O ponto de contacto para o conjunto de dados; A cobertura espacial do conjunto de dados; O período temporal que o conjunto de dados abrange; A data da última modificação do conjunto de dados; Os temas/categorias abordados por um conjunto de dados.”

Melhor prática 3: fornecer metadados estruturais, refere-se a fornecer metadados que descrevem o esquema e estrutura interna de uma distribuição, de forma a ajudar outras pessoas a consultar, explorar e perceber um conjunto de dados, bem como agentes de software poderão processar automaticamente as distribuições. “Os metadados estruturais legíveis por humanos normalmente fornecem as propriedades ou colunas do esquema de conjunto de dados. Os metadados estruturais legíveis por máquina estão disponíveis de

⁶ <https://www.w3.org/TR/html-rdfa/>

⁷ <https://www.w3.org/TR/json-ld/>

acordo com o formato de uma distribuição específica e podem ser fornecidos em documentos separados ou incorporados no documento.”

Melhor prática 4: fornecer informações sobre licença de dados, ou seja, fornecer um link ou cópia do contrato de licença dos dados. A licença dos dados é muito útil, pois fornece informação sobre as restrições de partilha e reutilização de um conjunto de dados. Esta licença pode estar especificada dentro dos metadados, ou num documento separado, mas vinculado a estes.

Melhor prática 5: fornecer informações de procedência de dados, diz respeito a fornecer informações sobre as origens dos dados e também de todas as alterações que estes já sofreram. A procedência é muito importante, uma vez que, permite aos consumidores de dados ter confiança nos dados, relativamente à sua integridade e credibilidade. Assim, com a procedência de dados, “os humanos conseguem saber a origem e histórico do conjunto de dados, enquanto os agentes de *software* são capazes de processar automaticamente informações de procedência”.

Melhor prática 6: fornecer informações de qualidade de dados. Indicar informações sobre a qualidade dos dados é muito importante no que diz respeito à qualidade das aplicações que fazem uso dos mesmos, podendo as aplicações ficar muito diferentes do propósito para que foram criadas. Com informações sobre a qualidade dos dados, o processo de seleção de conjuntos de dados fica mais facilitado, aumentando assim, as hipóteses de reutilização dos mesmos. Tanto como a qualidade deve ser documentada, os problemas relacionados com esta também devem ser declarados usando metadados. Ao disponibilizar informação de qualidade dos dados tanto os humanos como os agentes de *software* serão informados sobre a qualidade destes e assim, identificar se o conjunto é adequado para uma respetiva aplicação.

Melhor prática 7: fornecer um indicador de versão, que diz respeito a atribuir um número ou versão a um conjunto de dados, o que faz com que determinado conjunto de dados seja identificável. O indicador de versão de dados permite aos consumidores de dados identificar se os dados mudaram ao longo do tempo, mas também perceber com qual versão estão a trabalhar e se existe uma versão mais recente do mesmo conjunto de dados.

Melhor prática 8: fornecer histórico de versões, onde as alterações que foram feitas em cada versão são explicadas de forma a facilitar o processo de identificação de diferenças entre várias versões. Assim, humanos e agentes de *software* serão capazes de perceber como

o conjunto de dados muda de versão para versão e como duas versões específicas diferem. “Fornecer uma lista de versões publicadas e uma descrição para cada versão que explique como ela difere da versão anterior. Uma API pode expor um histórico de versão com uma única URL dedicada que recupera a versão mais recente do histórico completo.

Melhor prática 9: usar URIs persistentes como identificadores de conjuntos de dados. A utilização de identificadores comuns permite a identificação de conjuntos de dados e permite processos de comparação entre partes interessadas de forma confiável. Usar URIs é essencial para a gestão e reutilização dos dados. Os desenvolvedores de dados podem construir URIs no seu código, fazendo com que estes persistam e se refiram ao mesmo conjunto de dados ao longo do tempo. “Os conjuntos de dados ou informações sobre conjuntos de dados serão descobertos e citáveis ao longo do tempo, independentemente da disponibilidade ou formato dos dados.”.

Serviços de URIs persistentes são por exemplo: o Handle System, o DOI System, o Persistent Uniform Resource Locators (PURL), o URN System.

Melhor prática 10: usar URIs persistentes como identificadores dentro dos conjuntos de dados, ou seja, reutilizar os URIs entre conjuntos de dados e garantir que os seus identificadores podem ser referidos por outros conjuntos de dados. “Quando esses identificadores são URIs HTTP, eles podem ser pesquisados e mais dados descobertos”. Com isto, pretende-se que os dados se relacionem em toda a web, de forma a criar um espaço global de informações que são legíveis por humanos e por máquina.

Melhor prática 11: atribuir URIs a versões e séries de conjuntos de dados. Isto é útil para que humanos e agentes de *software* possam referir a versões específicas de determinado conjunto de dados e a conceitos como uma “série de conjuntos de dados” e a “versão mais recente”.

Melhor prática 12: usar formatos de dados padronizados legíveis por máquina, refere-se a disponibilizar dados que sejam processados por computador, de forma a minimizar as limitações na utilidade dos mesmos. Os dados só são úteis quando processados e transformados em informação. No que diz respeito ao formato dos dados, estes devem ser padronizados para evitar custos elevados e aumentar a eficiência, mas também para não perderem o sentido à medida que são transformados.

Melhor prática 13: usar representações de dados neutras ao local. Os dados tornam-se mais duradouros e menos abertos a más interpretações quando são legíveis por máquina e

não são específicos para alguma língua ou cultura específica. Em casos em que os dados já estão num formato específico de um local, deve-se fornecer metadados sobre o local onde esses valores são utilizados. Com os dados neutros, humanos e agentes de *software* podem interpretar o significado de strings com mais precisão.

Melhor prática 14: fornecer dados em vários formatos. Ao fornecer dados em vários formatos, reduz-se custos na transformação dos mesmos e reduz a possibilidade de introdução de erros nesse mesmo processo. Com os dados em diferentes formatos, aumenta o número de ferramentas e aplicações que processam os dados. Com isto, pretende-se conseguir o maior número de utilizadores de dados a usar dados, sem antes passar pelo processo de transformação dos mesmos para o formato desejado.

Melhor prática 15: reutilizar vocabulários, preferencialmente padronizados. Usar vocabulários partilhados por si só já facilita o consenso, aumenta a interoperabilidade e reduz as redundâncias. Para além disto, auxilia na comparação e no processamento automático de dados e metadados. “Organizações como o OGC, ISSO, W3C e WMO, bibliotecas e serviços de dados de pesquisa, etc. fornecem listas de códigos, terminologias e vocabulários de Dados Vinculados que podem ser usados por todos”.

Melhor prática 16: escolher o nível certo de formalização. Utilizar uma semântica formal, ajuda a transmitir especificações precisas e reduz o esforço em produzir e entender os dados, facilitando a sua reutilização. No entanto, se o vocabulário for demasiado simples, o editor de dados pode estar a “omitir alguns dados que fornecem informações importantes”.

Melhor prática 17: fornecer download em massa. Poder localizar os dados em massa pode ser útil, uma vez que se pode lidar com os dados como sendo um conjunto de dados, permitindo que estes sejam recuperados com uma única solicitação. Algumas formas de implementar esta prática passa por: “Para conjuntos de dados que existem inicialmente como vários arquivos, deve-se pré-processar uma cópia dos dados num único arquivo e tornar os dados acessíveis para download de um URI. Para conjuntos de dados maiores, o arquivo também pode ser comprimido; hospedar uma API que inclui a capacidade de recuperar um download em massa, além de consultas dinâmicas”.

Melhor prática 18: fornecer subconjuntos para grandes conjuntos de dados. Quando estamos perante um grande conjunto de dados, é importante dividir esse conjunto em subconjuntos. Assim, os utilizadores de dados podem fazer download apenas do subconjunto que precisam, em vez do conjunto inteiro de dados. O mesmo acontece com as

aplicações, se trabalharem com o conjunto completo de dados o seu desempenho irá diminuir, logo devem trabalhar com os subconjuntos que precisam, indo buscar novos sempre que necessitarem.

Melhor prática 19: usar negociação de conteúdo para servir dados disponíveis em vários formatos. Uma página HTML pode conter dados legíveis por humanos e máquina misturados, utilizando RDFa. “No entanto, como a Arquitetura da Web [WEBARCH]⁸ e DCAT[VOCAB-DCAT]⁹ deixam claro, um recurso, como um conjunto de dados, pode ter muitas representações. Os mesmos dados podem estar disponíveis como *JavaScript Object Notation (JSON)*, *Extensible Markup Language (XML)*, *Resource Description Framework (RDF)*, *Comma-Separated-Values (CSV)* e *HyperText Markup Language (HTML)*”. Estas várias representações de um conjunto de dados pode estar disponibilizado numa API, sendo que esta deverá estar disponibilizada no mesmo URL. Para isto, usa-se a negociação de conteúdo de forma que a representação desejada pelo utilizador seja retornada.

Melhor prática 20: fornecer acesso em tempo real. “Quando os dados são produzidos em tempo real e disponibilizados na web em tempo real ou quase em tempo real. Isso permite que as aplicações sejam capazes de aceder a dados temporalmente críticos em tempo real. Tempo real significa uma faixa de milissegundos a poucos segundos após a criação dos dados” e incentiva o desenvolvimento de aplicações em tempo real. A disponibilização de dados em tempo real, está dependente dos produtores de dados, que disponibilizem imediatamente os mesmos aos editores de dados.

Melhor prática 21: fornecer dados atualizados. Os dados devem ser disponibilizados perto do tempo da sua criação, ou do tempo da sua recolha. Para além disto, disponibilizar a frequência com que os mesmos são atualizados, aumenta a confiança dos utilizadores e o reaproveitamento dos dados.

Melhor prática 22: fornecer uma explicação para dados que não estão disponíveis. É importante publicar documentação online sobre dados que não se encontram disponíveis, de forma a identificar as lacunas de conhecimento e incentivar o uso dos dados que realmente estão disponíveis para a comunidade. “Dependendo do contexto máquina/humano, existem várias maneiras de indicar a indisponibilidade de dados. Os

⁸ <https://www.w3.org/TR/webarch/>

⁹ <https://www.w3.org/TR/vocab-dcat/>

editores de dados podem publicar um documento HTML que dá uma explicação legível para a indisponibilidade de dados. Do ponto de vista da interface da aplicação da máquina, códigos de status HTTP apropriados com mensagens personalizadas e legíveis por humanos podem ser usados.”

Melhor prática 23: disponibilizar dados através de uma API. Quando se está perante um grande conjunto de dados que é frequentemente atualizado ou altamente complexo poderá ser útil usar uma API para a publicação dos mesmos, pois permite maior flexibilidade para os consumidores de dados e permite a utilização de dados em tempo real. Para criar uma API não é necessariamente preciso iniciar uma do zero, pois é possível recorrer a plataformas de gestão de dados, de forma a utilizar uma já existente.

Melhor prática 24: usar padrões da web como base de APIs. Quando se cria uma API, é importante usar os padrões da web, de forma a aproveitar os seus pontos fortes. “*REST (REpresentational State Transfer)* é um estilo arquitetónico que, quando usado numa *APIWeb*, aproveita a arquitetura da própria Web. Outro aspecto da implementação a considerar é a criação de uma API hipermídia, que responde com links e dados”.

Melhor prática 25: fornecer documentação completa para a API. É importante disponibilizar na web informações sobre a API de maneira que os desenvolvedores de dados percebam a sua qualidade e utilidade. Para além disto, e não menos importante é necessário atualizar esta documentação, à medida que se adiciona recursos ou se faz alterações.

Melhor prática 26: evitar alterações drásticas na API. É necessário evitar alterações na API que façam com que o código do cliente deixe de funcionar, sendo necessário também comunicar todas as alterações que sejam feitas na API aos desenvolvedores, para que estes usufruam dos novos recursos e eventualmente tomarem medidas em caso de rutura. “Ao melhorar a API, concentre-se em adicionar novas chamadas ou novas opções em vez de mudar o funcionamento das chamadas existentes. Os clientes existentes podem ignorar tais mudanças e continuar a funcionar”.

Melhor prática 27: preservar identificadores. Se por alguma razão os dados necessitarem de ser removidos da web, é necessário preservar os respetivos identificadores para que o utilizador não seja direcionado para o código de resposta 404 (não encontrado). Para além disto, é preciso fornecer informação sobre o recurso arquivado. “Há dois cenários a considerar: o recurso foi excluído inteiramente e não está mais disponível por qualquer rota; o recurso foi arquivado e só está disponível através de uma solicitação ao arquivo.”

Melhor prática 28: avaliar a cobertura do conjunto de dados. Quando procedemos ao arquivamento de um conjunto de dados, é necessário avaliar a vinculação desse conjunto de dados a outros recursos já preservados e respectivos vocabulários utilizados.

Melhor prática 29: reunir *feedback* dos consumidores de dados. Receber *feedback* de dados, tem benefícios tanto para os editores de dados, como para os consumidores de dados. As vantagens que os editores de dados conseguem com o *feedback* dos consumidores, passa por estes conseguirem melhorar a integridade dos dados, para além de incentivar a partilha de novos dados. Assim, é importante fornecer aos consumidores de dados vários mecanismos de *feedback*, como por exemplo um “formulário de contacto, botões de classificação de qualidade de pontos e clique em dados ou uma caixa de comentários”.

Melhor prática 30: disponibilizar *feedback*. Ao disponibilizar *feedback* com os consumidores de dados, mostra que os editores têm interesse em mostrar que as preocupações destes estão a ser atendidas. Para além disto, a partilha de *feedback* também tem vantagens para os consumidores, uma vez que, conseguem perceber melhor possíveis problemas com os dados.

Melhor prática 31: enriquecer dados gerando novos dados. Enriquecer dados pode melhorar muito a processabilidade dos mesmos, principalmente quando se trata de dados não estruturados. “Em algumas circunstâncias, os valores perdidos podem ser preenchidos, e novos atributos e medidas podem ser adicionados a partir dos dados brutos existentes. Os conjuntos de dados também podem ser enriquecidos recolhendo resultados adicionais da mesma forma que os dados originais ou combinando os dados originais com outros conjuntos de dados. A publicação de conjuntos de dados mais completos pode aumentar a confiança, desde que seja feita de forma adequada e ética”.

Melhor prática 32: proporcionar apresentações complementares. É possível enriquecer os dados, ao apresentá-los de forma complementar, como visualizações, tabelas, aplicações web e resumos. Isto facilita a compreensão dos dados pelos humanos.

Melhor prática 33: fornecer *feedback* ao editor original. É importante informar o editor de dados original quando os seus dados estão a ser reutilizados, pois estes podem ter de fornecer estatísticas do uso dos dados. Para além disto, ajuda os editores originais a resolver erros ou até mesmo dar sugestões de melhoria. Uma forma de implementar esta prática passa por registar as informações do editor, do URI do conjunto de dados e data de quando recolhemos um novo conjunto de dados.

Melhor prática 34: cumprir os termos de licenciamento. Para isto, é preciso encontrar e seguir os requisitos de licenciamento do editor original dos dados para que estes confiem no trabalho. No caso dos utilizadores de dados, eles podem definir a licença dos trabalhos derivados de forma adequada. A forma de pôr isto em prática passa por “ler a licença original e cumprir todas as exigências. Se a licença solicitar o licenciamento específico de obras derivadas, é necessário escolher uma licença para ser compatível com esse requisito. Se nenhuma licença for dada, é necessário entrar em contacto com o editor original e perguntar qual é a licença.”

Melhor prática 35: citar a publicação original. Identificar a fonte dos dados em metadados, fornece confiabilidade aos mesmos, uma vez que, o utilizador pode tirar as suas próprias conclusões quanto à reputação da fonte e permite transmitir confiança como reutilizador. Para além disto, citar a fonte de dados aumenta o reconhecimento dos editores, incentivando-os a publicar mais dados. “É possível apresentar a citação à fonte original numa interface de utilizador fornecendo texto bibliográfico e um link de trabalho (Lóscio et al., 2017).

Os principais desafios a que as boas práticas respondem são: a disponibilização de metadados, a licença dos dados, proveniência e qualidade, versionamento, identificação, formato, vocabulários, acesso, preservação, *feedback*, enriquecimento e republicação de dados (Lóscio et al., 2017).

Para além de responder aos vários desafios, as DWBP trazem inúmeros benefícios. De seguida são apresentados os benefícios e as boas práticas correspondentes (Lóscio et al., 2017):

a) Compreensão

- fornecer metadados;
- fornecer metadados descritivos;
- fornecer metadados estruturais;
- fornecer informações de proveniência de dados;
- usar representações de dados neutras em termos de localidade;
- reutilizar vocabulários, de preferência padronizados;
- escolher o nível de formalização correto;
- obter *feedback* de consumidores de dados;
- enriquecer os dados gerando novos dados;
- fornecer apresentações complementares;

b) Capacidade de processamento

- fornecer metadados;

- fornecer metadados estruturais;
 - usar formatos de dados padronizados legíveis por máquina;
 - fornecer dados em vários formatos;
 - reutilizar vocabulários, de preferência padronizados;
 - fornecer subconjuntos para grandes conjuntos de dados;
 - disponibilizar dados através de uma API;
 - usar web standards como base das APIs;
 - enriquecer os dados gerando novos dados;
- c) Reutilização
- todas as DWBP;
- d) Capacidade de descoberta
- fornecer metadados;
 - fornecer metadados descritivos;
 - usar URIs persistentes como identificadores de conjuntos de dados;
 - usar URIs persistentes como identificadores em conjuntos de dados;
 - atribuir URIs a versões e séries de conjuntos de dados;
 - usar web standards como base das APIs;
 - citar a publicação original;
- e) Confiança
- fornecer informações de licença de dados;
 - fornecer informações de proveniência de dados;
 - fornecer informações de qualidade de dados;
 - fornecer um indicador de versão;
 - fornecer o histórico da versão;
 - atribuir URIs a versões e séries de conjuntos de dados;
 - reutilizar vocabulários, de preferência padronizados;
 - fornecer uma explicação para os dados que não estão disponíveis;
 - fornecer documentação completa para a API;
 - evitar interromper alterações na API;
 - preservar identificadores;
 - avaliar a cobertura do conjunto de dados;
 - obter *feedback* de consumidores de dados;
 - disponibilizar *feedback*;
 - enriquecer os dados gerando novos dados;
 - fornecer apresentações complementares;
 - fornecer *feedback* ao editor original;
 - seguir os termos de licenciamento;
 - citar a publicação original;
- f) *Linkability*
- usar URIs persistentes como identificadores de conjuntos de dados;
 - usar URIs persistentes como identificadores em conjuntos de dados;
 - fornecer subconjuntos para grandes conjuntos de dados;
 - usar web standards como base das APIs;
- g) Acesso

- fornecer download em massa;
 - fornecer subconjuntos para grandes conjuntos de dados;
 - usar a negociação de conteúdo para servir dados disponíveis em vários formatos;
 - fornecer acesso em tempo real;
 - fornecer dados atualizados;
 - disponibilizar dados através de uma API;
 - usar web standards como base das APIs;
 - fornecer apresentações complementares;
- h) Interoperabilidade
- usar URIs persistentes como identificadores de conjuntos de dados;
 - usar URIs persistentes como identificadores em conjuntos de dados;
 - reutilizar vocabulários, de preferência padronizados;
 - escolher o nível de formalização correto;
 - disponibilizar dados através de uma API;
 - usar web standards como base das APIs;
 - evitar interromper alterações na API; e
 - fornecer *feedback* ao editor original.

O próximo capítulo refere-se ao enquadramento metodológico adotado para a análise de qualidade de conjuntos de dados referentes às categorias Governo e Setor Público, Transportes e Educação, Cultura e Desporto do EDP.

3. ENQUADRAMENTO METODOLÓGICO

De forma a compreender o desenvolvimento do presente estudo é importante perceber o enquadramento metodológico e sob qual ótica se realizou a pesquisa e como esta foi executada para atingir os seus objetivos e resultados finais.

Assim, quanto ao objetivo da pesquisa, este estudo caracteriza-se como sendo uma pesquisa exploratória. Este tipo de pesquisa ocorre quando não há muito conhecimento sobre o tema a ser estudado e tem como objetivo aprofundar o tema de forma a que este fique mais claro, ou construir questões importantes para a pesquisa (Raupp & Beuren, 2013). Uma característica importante deste tipo de pesquisa é que esta consiste no “[...] aprofundamento de conceitos preliminares sobre determinada temática não contemplada de modo satisfatório anteriormente” (Raupp & Beuren, 2013, p. 5, tradução própria). Este tipo de pesquisa constitui-se como o primeiro passo no campo científico, possibilitando a realização de outros tipos de pesquisas relacionadas com o mesmo tema (Raupp & Beuren, 2013).

Relativamente à abordagem metodológica adotada para este estudo, optou-se pela quali-quantitativa. Na abordagem quantitativa, a explicação de resultados é feita através da recolha de dados numéricos que serão posteriormente analisados através de métodos matemáticos (Paschoarelli et al., 2015). Logo, este tipo de abordagem procura precisão nos resultados, de forma a evitar dúvidas na análise dos mesmos (Paschoarelli et al., 2015).

A abordagem qualitativa diferencia-se da abordagem quantitativa pela forma como os dados são recolhidos, tratados e analisados. Este tipo de abordagem não utiliza instrumentos estatísticos para a análise do problema e a recolha de dados ocorre através do contacto direto entre o pesquisador e o objeto em estudo (Paschoarelli et al., 2015).

A abordagem quali-quantitativa cobre os aspetos fortes da abordagem qualitativa e quantitativa, procurando assim, alcançar uma maior compreensão dos problemas em estudo. Esse tipo de abordagem é adequada para minimizar a subjetividade e aproximar o pesquisador do objeto estudado, dando mais credibilidade aos dados (Paschoarelli et al., 2015).

No que diz respeito ao resultado da pesquisa, este estudo caracteriza-se como pesquisa aplicada, uma vez que, tem como objetivo produzir conhecimento científico para aplicação prática (For et al., 2009). Para além de produzir conhecimento científico, este tipo de pesquisa também gera novos processos tecnológicos e novos produtos (For et al., 2009).

Os procedimentos técnicos explicam a forma pela qual se conduz o estudo e como se obtém os dados. Assim, os procedimentos técnicos utilizados foram: a pesquisa bibliográfica, ou de fontes secundárias, que diz respeito a toda a bibliografia que já foi tornada pública em relação ao tema estudado (Lakatos & Marconi, 2003) e pode ir desde publicações, revistas, jornais, livros, pesquisas, teses, e até mesmo, rádio, gravações, tendo como objetivo o contacto direto entre pesquisador e tudo o que é escrito, dito ou filmado sobre o tema; a pesquisa documental, onde a recolha dos dados está limitada a documentos, podendo estes estar escritos, ou não, sendo denominada como fonte primária; e o estudo de caso, que se caracteriza como o estudo concentrado em um único caso (Raupp & Beuren, 2013) de forma a aprofundar o conhecimento sobre um caso específico, o European Data Portal.

Como ferramenta de apoio a organização dos documentos utilizados para fundamentação teórica é adotado o Mendeley^(TM).

3.1 Procedimentos Metodológicos

No European Data Portal, os dados estão classificados em quatro grupos de classificação: Excelente, Bom, Suficiente e Mau. A classificação de cada conjunto de dados é a pontuação geral que um conjunto de dados obtém, a partir da soma dos pontos de cada dimensão.

As dimensões avaliadas no Portal são as seguintes: Pesquisabilidade (0 a 100 pontos), Acessibilidade (0 a 100 pontos), Interoperabilidade (0 a 110 pontos), Reusabilidade (0 a 75 pontos) e Contextualidade (0 a 20 pontos) (Figura 1) (<https://data.europa.eu/mqa/methodology>).

| Dimensão | Pontos máximos |
|--------------------|----------------|
| Pesquisabilidade | 100 |
| Acessibilidade | 100 |
| Interoperabilidade | 110 |
| Reutilizabilidade | 75 |
| Contextualidade | 20 |
| Soma | 405 |

Figura 1 - Pontos máximos para cada dimensão (reproduzido de Data.europa.eu, 2021).

Na dimensão Pesquisabilidade, observam-se métricas que medem como as pessoas e máquinas encontram o conjunto de dados. Esta dimensão avalia os seguintes indicadores: utilização de palavras-chave, categorias, pesquisa espacial e pesquisa temporal (Data.europa.eu, 2021).

A dimensão Acessibilidade pretende medir se o acesso aos dados referenciados pelas distribuições está garantido. Nesta dimensão, encontram-se os seguintes indicadores: acessibilidade da URL de acesso, URL de transferência e acessibilidade da URL de transferência.

A dimensão Interoperabilidade permite verificar se uma distribuição é considerada interoperável. Nesta dimensão encontram-se os seguintes indicadores: formato, tipo de suporte, formato/tipo de suporte do vocabulário, não proprietário, legível por máquina e conformidade com o perfil de aplicação DCAT.

A dimensão Reusabilidade descreve medidas que são utilizadas para verificar a reusabilidade dos dados. Esta dimensão inclui os seguintes indicadores: informações constantes da licença, vocabulário da licença, restrições de acesso, vocabulário das restrições de acesso, ponto de contacto e publicador.

A dimensão Contextualidade apresenta propriedades que podem fornecer mais contexto ao utilizador. Nesta dimensão, encontram-se os seguintes indicadores: direitos, tamanho do ficheiro, data de emissão e data de alteração.

Informações adicionais sobre cada um destes indicadores podem ser consultadas em <https://www.europeandataportal.eu/mqa/methodology?locale=pt>.

A soma dos valores de todos os indicadores de cada dimensão resulta na pontuação de cada conjunto de dados. Conforme os pontos obtidos, os conjuntos de dados são divididos em quatro classificações, como já foi referido anteriormente. A classificação de mau, inclui conjuntos de dados cuja pontuação se encontra entre 0 a 120 pontos. A classificação de suficiente representa conjuntos de dados que tiveram uma pontuação entre 121 e 220 pontos. A classificação de bom diz respeito a conjuntos de dados que tiveram uma pontuação que corresponde entre 221 a 350 pontos, e a classificação de excelente são aqueles conjuntos de dados que tiveram as melhores pontuações, sendo que estas podem variar entre os 351 pontos até à pontuação máxima de 405 pontos (Figura 2) (Data.europa.eu, 2021).

| Notação | Escala de pontos |
|------------|------------------|
| Excelente | 351 - 405 |
| Bom | 221 - 350 |
| Suficiente | 121 - 220 |
| Mau | 0 - 120 |

Figura 2 - Escala de Pontos para cada classificação (reproduzido de Data.europa.eu, 2021).

3.2 Estudo Exploratório

No estudo exploratório serão analisados os conjuntos de dados do EDP relativos ao Governo e Setor Público, Transportes e Educação, Cultura e Desporto. À data de 04/Jan/2021, na categoria Governo e Setor Público no European Data Portal, encontram-se 59.163 conjuntos de dados (Figura 3).

Na categoria de Transportes, à data 04/Jan/2021, identificaram-se 40.423 conjuntos de dados (Figura 3). Por último, na categoria de Educação, Cultura e Desporto, à data de 04/Jan/2021, encontraram-se 10.992 conjuntos de dados (Figura 3).



Figura 3 - Conjuntos de dados encontrados para a categoria Governo e Setor Público, Transportes, Educação, Cultura e Desporto (reproduzidos de Data.europa.eu, 2021).

Como não é possível analisar todos os conjuntos de dados manualmente em tempo útil, tornou-se necessário recorrer a amostragem. Iniciou-se, por isso, por definir critérios de amostragem e por aplicá-los num estudo exploratório. Foi efetuada uma amostragem sistemática dos conjuntos de dados com a classificação¹⁰ de Excelente e Bom. Para esse fim, foi desenvolvido o algoritmo a seguir:

```
INICIO
  LER (classificação Excelente/Bom+);
  SE ordenado_last_modified==verdade ENTAO
    Analisar os 20 primeiros de cada categoria ();
  SENA0
    Ordenar();
    Analisar os 20 primeiros de cada categoria ();
  FIM SE
FIM
```

O estudo exploratório incidiu sobre os primeiros 20 conjuntos de dados encontrados em cada categoria e visou:

- a) verificar a adequabilidade do algoritmo de constituição da amostra, em particular identificar desvios significativos da amostra relativamente à população (por exemplo, se os conjuntos de dados tiverem todos o mesmo idioma, ou tiverem origem no mesmo país, será necessário fazer alterações). O objetivo não é obter uma amostra representativa, mas também não se pretende obter uma amostra com desvios excessivos relativamente à população; e
- b) obter os primeiros resultados do estudo, identificar potenciais dificuldades de implementação ou desvios ao que estava inicialmente planeado. A prossecução deste objetivo permitir-nos-á realizar os devidos ajustes em tempo hábil, robustecendo os resultados futuros.

Assim, na categoria de Governo e Setor Público, com a classificação de excelente e ordenado pela última alteração foram encontrados 49 conjuntos de dados. Destes 49, foram analisados os 20 primeiros (a pesquisa para análise foi realizada no dia 06/Jan/2021) (Figura 4).

¹⁰ Esta classificação dos conjuntos de dados oscila a depender da avaliação do EDP realizada regularmente.

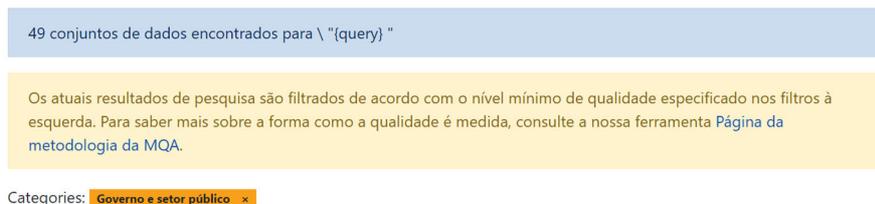


Figura 4 - Conjuntos de dados encontrados para a categoria de Governo e Setor Público, com a classificação de excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Na categoria de Transportes com a classificação de excelente e ordenado pela última alteração não foi encontrado nenhum conjunto de dados (a pesquisa para análise foi realizada no dia 07/Jan/2021) (Figura 5).

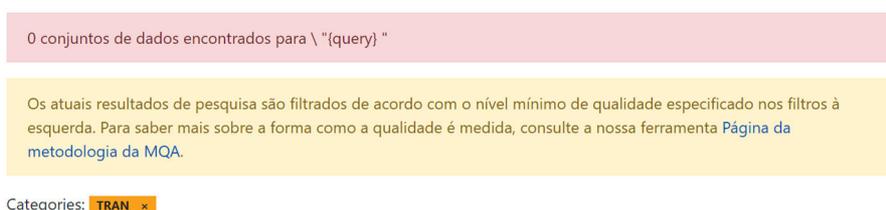


Figura 5 - Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Como não foi encontrado nenhum conjunto de dados na classificação de excelente, foi necessário utilizar a classificação de Bom+. Com esta classificação e ordenado pelo último modificado, foram encontrados 2.128 conjuntos de dados, dos quais analisou-se os 20 primeiros desta categoria (a pesquisa para análise foi realizada no dia 07/Jan/2021) (Figura 6).

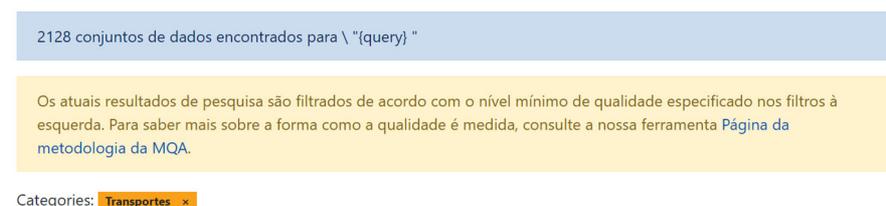


Figura 6 - Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de Bom+ e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Na categoria Educação, Cultura e Desporto com a classificação de excelente e ordenado pela última alteração, apenas foi encontrado um conjunto de dados (a pesquisa para análise foi realizada no dia 08/Jan/2021) (Figura 7).

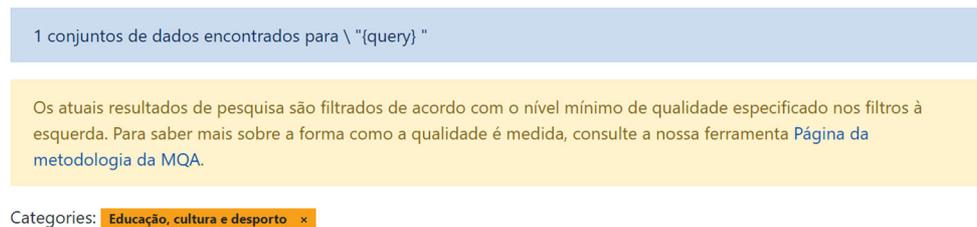


Figura 7 - Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Como apenas foi encontrado um resultado na classificação de excelente, levou a que a classificação para análise passe de excelente para Bom+. Assim, com a classificação Bom+ e ordenado pelo último modificado, foram encontrados 1.978 conjuntos de dados, onde foram analisados os 19 primeiros, mais o único conjunto de dados encontrado, marcado com a classificação de excelente (A pesquisa para análise foi realizada no dia 08/Jan/2021) (Figura 8).

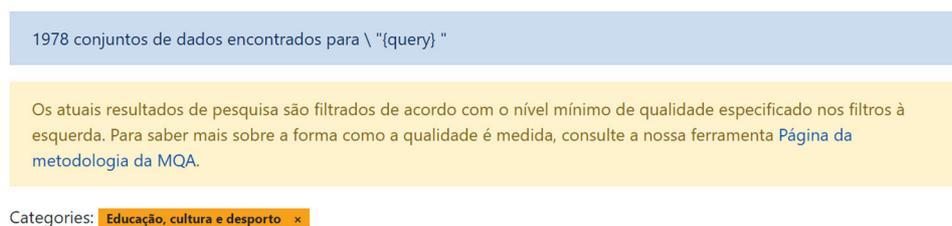


Figura 8 - Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Bom+ e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

3.3 Estudo Final

Depois de realizado o estudo exploratório, verificou-se que a técnica de amostragem utilizada, ou seja, a análise dos 20 primeiros conjuntos de dados de cada categoria, filtrados por Excelente e Bom+, não foi a melhor, uma vez que, foram encontrados muitos conjuntos de dados pertencentes ao mesmo país, onde os dados se referiam ao mesmo tópico, sendo o que alterava, era o ano a que se referiam.

Logo, era necessário encontrar uma técnica de amostragem que fosse estatisticamente significativa, de forma a conseguir analisar todos os dados e não só aqueles já identificados. Desta forma, a nova técnica de amostragem utilizada é a amostragem por conveniência. Com os conjuntos de dados ordenados pelo último modificado e filtrados pela classificação de Excelente e Bom+, percorre-se a listagem que é apresentada. Ao percorrer esta listagem, é

possível detetar se um conjunto de dados pertence à mesma entidade, através do nome do conjunto de dados e se os dados são idênticos, apenas alterando o ano a que se referem, por exemplo.

No estudo final foi utilizado um novo algoritmo de forma a analisar os conjuntos de dados de acordo com a técnica de amostragem de conveniência. O algoritmo utilizado é apresentado abaixo:

```
START
    Sort Excellent and Good+ datasets by EDP rate;
    Second sort the datasets by coverage date;
    Count=0;
    While (count <50)
        If new dataset similar to dataset previously analyzed THEN
            remove dataset from the list;
        else
            add dataset to the sample;
            remove dataset from the list;
            count= count+1;
    End_while
END
```

Assim, o método passa por analisar 50 conjuntos de dados, desde que não sejam idênticos, no sentido de corresponder ao mesmo tema apenas com datas diferentes, independentemente da fonte. Quando isto acontecer, passa-se à frente até atingir os 50 conjuntos de dados.

Com a classificação de Excelente e ordenado pela última modificação (à data de 06/Mai/2021), não foi encontrado nenhum conjunto de dados na categoria de Governo e Setor Público (Figura 9).

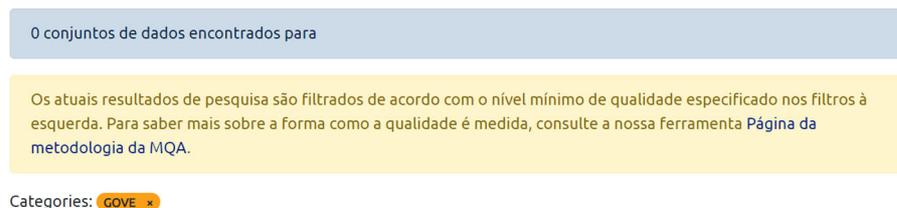


Figura 9 - Conjuntos de dados encontrados para a categoria de governo e setor público, com a classificação de excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Como na classificação de excelente não foram encontrados nenhum conjunto de dados, foi necessário passar para a classificação de Bom+, onde já foram encontrados 534 conjuntos de dados. Destes, analisaram-se os 50 primeiros ordenados pela última alteração, desde que não pertençam à mesma entidade, com o mesmo título, apenas com os anos a que se referem diferentes (Figura 10).

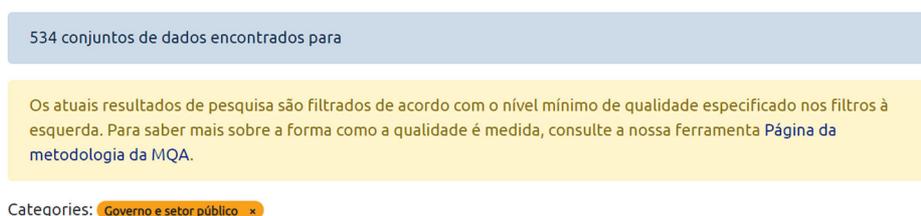


Figura 10 - Conjuntos de dados encontrados para a categoria de governo e setor público, com a classificação de Bom+ e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Na categoria Transportes, com a classificação de Excelente e ordenado pelo último modificado não foram encontrados conjuntos de dados (análise realizada à data de 11/Mai/2021) (Figura 11).

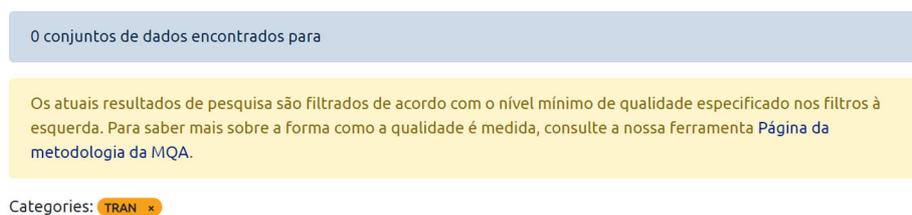


Figura 11 - Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Assim sendo, como na classificação de Excelente não foi encontrado nenhum conjunto de dados, foi necessário passar para a classificação de Bom+. Nesta classificação, à data de 11/Mai/2021, foram encontrados 2.563 conjuntos de dados, dos quais foram analisados os 50 primeiros de acordo com os critérios de exclusão definidos (Figura 12).

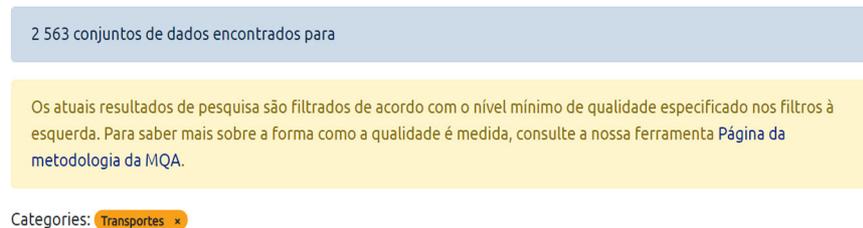


Figura 12 - Conjuntos de dados encontrados para a categoria de Transportes, com a classificação de Bom+ e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Com a classificação de Excelente e ordenado pelo último modificado, a categoria de Educação, Cultura e Desporto, à data de 14/Mai/2021, não apresenta nenhum conjunto de dados (Figura 13).

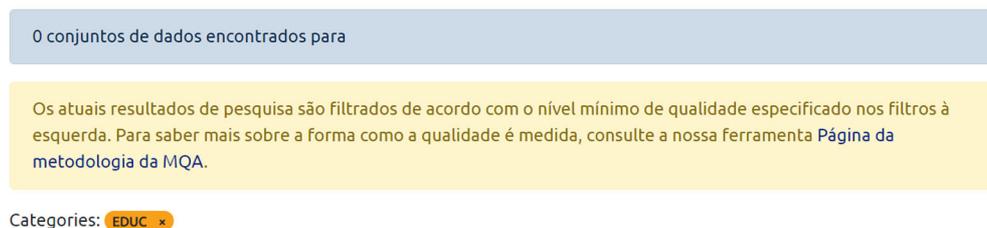


Figura 13 - Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Excelente e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

Como não foram encontrados conjuntos de dados nesta classificação, foi necessário passar para a classificação de Bom+, onde foram encontrados 4.343 conjuntos de dados. Ordenados pelo último modificado, foram analisados os 50 primeiros, de acordo com a técnica de amostragem estabelecida (análise realizada à data de 14/Mai/2021) (Figura 14).

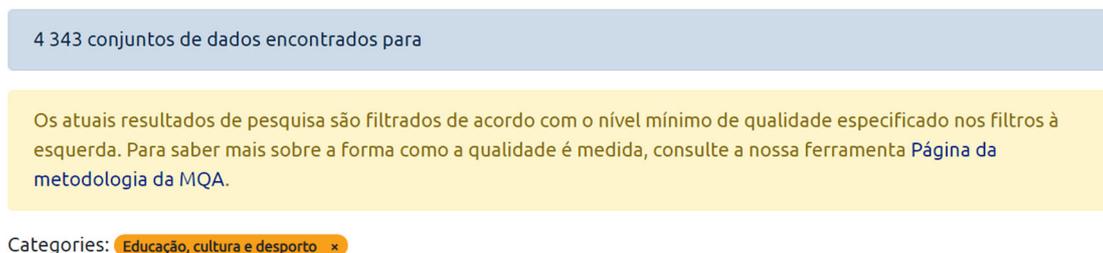


Figura 14 - Conjuntos de dados encontrados para a categoria de Educação, Cultura e Desporto, com a classificação de Bom+ e ordenado pela última alteração (reproduzido de Data.europa.eu, 2021).

3.3.1 Critérios para a Análise dos Conjuntos de Dados

Para analisar os conjuntos de dados de cada categoria, como já foi referido, recorreu-se às boas práticas de Dados na Web, da W3C. Assim, à luz das DWBP analisaram-se os metadados das categorias do Governo e Setor Público, Transportes e Educação, Cultura e Desporto.

A análise de cada conjunto de dados, de forma a verificar se cada uma das boas práticas estavam presentes, foi realizada com base nas informações referidas na página principal de cada um dos conjuntos, mas também na análise dos dados por meio do Turtle, onde é possível visualizar os dados, assim como os metadados empregados.

Para cada boa prática, foram definidos os critérios que permitem verificar se ela está contemplada por um determinado conjunto de dados, ou não. Ao analisar os conjuntos de dados segundo as DWBP, foi colocado o código “sim”, ou “não”, ou “não identificada” ou “não aplicada”. Para além disto, para cada boa prática criou-se o campo de observações, de forma a registar algum aspeto atípico encontrado na análise dos dados.

Em seguida, são apresentados para cada boa prática os respetivos critérios:

Boa prática 1– Fornecer metadados. Para esta boa prática, coloca-se em todos os conjuntos de dados que esta estava presente, com o código “sim”, uma vez que o European Data Portal obriga ao fornecimento de metadados.

Boa prática 2– Fornecer metadados descritivos. Se o conjunto de dados apresentar uma descrição sobre os dados, palavras-chave, título, publicador, datas e formato então coloca-se que esta boa prática se observa, se não, coloca-se que “não” e acrescenta-se no campo apropriado de observações as propriedades tidas como essenciais que não estão presentes.

Boa prática 3– Fornecer metadados estruturais. Para esta boa prática é necessário abrir cada conjunto de dados e ver se contém metadados estruturais. Se tiver coloca-se “sim”, se não tiver, coloca-se “não”.

Boa prática 4– Fornecer informação de licença de dados. Se o conjunto de dados tiver uma licença e o tipo de licença for especificado, então coloca-se "sim", caso contrário coloca-se "sim", acrescenta-se informações relevantes no campo de observações. Se não tiver uma licença, então coloca-se "não".

Boa prática 5– Fornecer informação sobre a proveniência dos dados. Para verificar se esta BP foi adotada é necessário verificar se estão presentes as seguintes propriedades:

dct:creator/dct:issued, ou dct:publisher/dct:issued, ou dct:creator, dct:publisher e dct:issued, ou prov:actedOnBehalfOf. Se isto se verificar, coloca-se “sim”; se não, “não”.

Boa prática 6– Fornecer informação de qualidade de dados. Se o conjunto de dados tiver a propriedade dqv:QualityMeasurement, colocar "sim"; caso contrário, colocar "não".

Boa prática 7– Fornecer indicador de versão. Classificada em BP7H, informação legível por humanos, e BP7C, informação legível por computador. Para a BP7H, se o conjunto de dados apresentar informação de versão nas informações adicionais do conjunto de dados, coloca “sim”; se não, coloca “não”. Para a BP7C, se a propriedade owl:versionInfo estiver presente, colocar “sim”; se não, “não”.

Boa prática 8– Fornecer histórico de versão. Classificada em BP8H, informação legível por humanos, e BP8C, informação legível por computador. Para a BP8H, se o conjunto de dados apresentar a data de criação na página principal, coloca-se “sim”, se não, coloca-se “não”. Para a BP8C, se o conjunto de dados apresentar as propriedades dct:isVersionOf , owl:versionInfo, pav:version, pav:previousVersion ou dct:hasVersion coloca-se “sim”, se não, coloca-se “não”.

Boa prática 9– Usar URIs como identificadores persistentes de conjuntos de dados. Para esta boa prática, se o conjunto de dados utiliza serviços conhecidos como: URN System, Handle System, DOI System, ARK System, Persistent Uniform Resource Locators (PURL), Electronic Identifier Serial Publications (EISPs), International eBook Identifier Numbers (IEINs), Identificadores de recursos extensíveis (XRI), Links magnéticos – magnet, Virtual International Authority File (VIAF), International Standard Name Identifier (ISNI) e International Standard Name Identifier (ISNI). Quando observado um dos identificadores supracitados, coloca-se sim, do contrário, não.

Boa prática 10- Use URIs como identificadores dentro de conjuntos de dados. Para esta boa prática, foi necessário verificar se dentro de cada conjunto de dados, referências a países, regiões, organizações e pessoas estão referidas por URIs. Para isto, verifica-se se propriedades como dct:creator, dct:publisher, dct:location, dct:spacial, dct:subject, dct:licence e dct:contributor estão referidos por URI. Se estiverem, coloca-se “sim”; se não, coloca-se “não”.

Boa prática 11– Usar URIs para as versões e séries de conjuntos de dados. Verifica-se se está atribuído um URI a cada versão. Se estiver coloca-se “sim”; se não, coloca-se “não”.

Boa prática 12– Fornecer formatos legíveis por máquina. Se o conjunto de dados, na secção *Linked data* apresentar as opções de visualização em formato RDF/XML, Turtle, Notation3, N-Triples e JSON-LD colocar “sim”, se não, colocar “não”. Se o conjunto de dados for em formato PDF colocar “não”.

Boa prática 13– Usar representações de dados neutros. Classificada em BP13H, informação legível por humanos, e BP13C, informação legível por computador. Para a BP13H se o conjunto de dados apresentar a localização, coordenadas ou indicação de conformidade com, coloca-se “sim”; se não, coloca-se “não”. Para a BP13C se o conjunto de dados apresentar as propriedades: *dct:conformsTo*, *dct:language*, *dct:location* e *dct:spacial*, coloca-se “sim”; se não, coloca-se “não”.

Boa prática 14– Fornecer dados em vários formatos. Observa-se se os conjuntos de dados apresentam distribuições em vários formatos. Se apresentar, coloca-se “sim”; se não, coloca-se “não”.

Boa prática 15– Reutilizar vocabulário, preferencialmente padronizados. “Os vocabulários definem os conceitos e as relações (também referidos como "termos" ou "atributos") utilizados para descrever e representar uma área de interesse. São utilizados para classificar os termos que podem ser utilizados numa determinada aplicação" (Lóscio, 2017). Os vocabulários são adotados para descrever os metadados, tais como *dcat*, que incorpora vocabulários pré-existentes como Dublin Core, Foaf, *vCard*; e para representar os dados, com vocabulários como Unesco, Eurovoc thesaurus, EuroSciVoc taxonomy, e lista de autoridades. Todos os conjuntos de dados analisados, no EDP apresentam a propriedade *dcat:theme* (catálogo de autoridades). Dessa forma, foi verificado apenas se eram utilizados vocabulários controlados a exemplo do Eurovoc para a propriedade *dct:subject*. Nesse caso, coloca-se “sim” se apresentar a propriedade *dct:subject*, ou coloca-se “não” se não apresentar.

Boa prática 16– Escolher o nível certo de formalização. Para esta melhor prática, se o conjunto de dados utilizar vocabulários apropriados, como Dublin Core e Schema.org, para descrever, coloca-se "sim"; se não, coloca-se "não".

Boa prática 17- Fornecer *download* em massa. Se o conjunto de dados pode ser descarregado de uma só vez, colocar "sim"; caso contrário, colocar "não".

Boa prática 18– Fornecer subconjuntos para grandes conjuntos de dados. Esta boa prática só se aplica a grandes conjuntos de dados. No EDP estes já se encontram divididos.

Boa prática 19– Usar negociação de conteúdo. Se o conjunto de dados apresentar as várias formas de representação disponíveis em *Linked Data*, colocar “sim”, se não, colocar “não”.

Boa prática 20– Fornecer acesso em tempo real. Não aplicada.

Boa prática 21– Fornecer dados atualizados. Para esta boa prática, verifica-se se o conjunto de dados apresenta a propriedade *dct:accrualPeriodicity*, se apresentar coloca-se “sim”, se não, coloca-se “não”.

Boa prática 22– Fornecer explicação para dados que não estão disponíveis. Classificada em BP22H, informação legível por humanos e BP22C, informação legível por computador. Para BP22H, se os conjuntos de dados estiverem acompanhados por um documento HTML com informações sobre os dados lá referidos, mas não disponíveis, e com explicações sobre a razão da falta e formas de acesso, colocar "sim"; caso contrário, colocar "não". Não foi verificada a BP22C.

Boa prática 23– Disponibilizar dados através de uma API. A EDP possibilita a distribuição de conjuntos de dados por API, mas como não dependem dos conjuntos de dados, não foi analisada.

Boa prática 24– Usar padrões da Web como base de APIs. Esta BP não depende dos conjuntos de dados, portanto não foram analisadas.

Boa prática 25– Disponibilizar documentação completa da API. Esta BP não depende dos conjuntos de dados, portanto não foi analisada.

Boa prática 26– Evitar quebrar alterações que afetem o funcionamento da API. Esta BP não depende dos conjuntos de dados, portanto não foi analisada.

Boa prática 27– Preservar identificadores. Esta também não é aplicável, uma vez que não se analisa conjuntos de dados removidos, só se verifica conjuntos de dados disponíveis. Embora se saiba que alguns *data providers* não tiveram este cuidado, isto está fora do âmbito deste trabalho.

Boa prática 28– Avaliar a cobertura do conjunto de dados. Não aplicada, uma vez que está relacionada com a preservação do arquivo e não com a disponibilização. Neste trabalho apenas se trabalhou com a disponibilização.

Boa prática 29– Reunir *feedback* dos consumidores de dados. Verifica-se se existe mecanismo de *feedback* dos consumidores de dados no EDP, como o *e-mail* ou outro canal de comunicação para entrar em contacto. Se tiver, coloca-se “sim”, se não, “não”.

Boa prática 30– Disponibilizar *feedback*. Se o conjunto de dados apresentar a propriedade `rdfs:comment`, colocar “sim”; se não, colocar “não”.

Boa prática 31– Enriquecer dados por meio de geração de novos dados. Para esta melhor prática, se o conjunto de dados original tiver sido enriquecido, se os dados resultantes do enriquecimento forem identificados e se os dados originais antes do enriquecimento ainda se mantiverem, colocar “sim”; caso contrário, colocar “não”. Esta BP não foi analisada por ter pouca relevância para o contexto deste estudo, uma vez que o objectivo é analisar os conjuntos de dados em si, e não os dados em si.

Boa prática 32– Proporcionar apresentações complementares. Se o conjunto de dados mostrar apresentações complementares tais como um gráfico, colocar “sim”; caso contrário, colocar “não”.

Boa prática 33– Fornecer *feedback* ao editor original. A conformidade com esta BP está fora do âmbito deste estudo uma vez que não temos acesso à comunicação entre a EDP e os seus fornecedores de dados.

Boa prática 34– Seguir os termos de licenciamento. Para analisar esta boa prática, é necessário verificar a licença dos conjuntos de dados de acordo com o termo apresentado. Como o EDP recolhe os dados com o mesmo tipo de licença fornecida na origem, foi colocado “sim” para todos os conjuntos de dados (<https://data.europa.eu/pt/faq>) (European Data Portal, 2020).

Boa prática 35– Citar a publicação original. É necessário verificar quais os que citam e não citam a publicação original.

3.4 Procedimentos Adotados

No estudo exploratório e no estudo final foram utilizadas as folhas do Excel, para analisar os conjuntos de dados de cada categoria. Nas colunas encontram-se as boas práticas e nas linhas o identificador de cada conjunto de dados. A análise do estudo exploratório foi realizada com o código: “Sim”, “Não”, “Não Identificado” e “Não Aplicado” para cada uma das DWBP, enquanto a análise de cada DWBP no estudo final foi realizada com o código: “Sim”, “Não” e “Não Aplicado”.

Como referido anteriormente, o EDP organiza os conjuntos de dados das várias categorias por diferentes classificações. Para cada indicador avaliado pelo Portal, está associado um peso. Se o conjunto de dados cumprir com um determinado indicador é-lhe atribuído esse peso correspondente, sendo que no final é somado todos pesos para saber qual a pontuação do conjunto de dados, sendo-lhe atribuído conforme a pontuação atingida, uma classificação.

Assim, de forma a pontuar cada conjunto de dados analisado à luz das DWBP, foi atribuído a cada BP uma pontuação, como o EDP pontuou cada indicador, para classificar cada conjunto de dados.

Como referido anteriormente no tópico 2.6, a utilização das boas práticas traz certos benefícios. Para cada benefício foi atribuído um único peso, sendo que a pontuação de cada BP está dependente do número de benefícios que esta obtém. Ou seja, a pontuação de cada BP está relacionada com o número de benefícios obtidos que esta traz na sua utilização.

A aplicação das DWBP permite a obtenção de, no máximo, um total de oito benefícios, sendo que, se uma BP tiver os 8 benefícios esta apresenta 100% dos benefícios. No entanto, como é possível observar na Tabela 1, o máximo de benefícios que uma BP consegue atingir é 5. Logo, para ter um peso igual para cada benefício, dividiu-se os 100% pelos cinco benefícios, o que resulta que cada benefício tenha um peso igual de 20 pontos.

Na Tabela 1, as BPs não aplicadas não estão incluídas, de forma a não interferir com a pontuação de cada conjunto de dados. As BPs não aplicadas são: as BP18, BP20, BP23, BP24, BP25, BP26, BP27, BP28, BP31 e BP33.

Tabela 1 - Pontuação de cada BP

| Dimensão | Subdimensão | Benefício | Peso |
|--------------------------|--|-----------|------|
| Metadados | Fornecer metadados | | 80 |
| | Fornecer metadados descritivos | | 60 |
| | Fornecer metadados estruturais | | 60 |
| Licença de Dados | Fornecer informações sobre a licença de dados | | 40 |
| Proveniência e Qualidade | Fornecer informações de proveniência dos dados | | 60 |
| | Fornecer informações de qualidade dos dados | | 40 |
| Versionamento de Dados | Fornecer indicador de versão (versão legível por humano) | | 40 |

Tabela 2 - Pontuação de cada BP (Continuação)

| Dimensão | Subdimensão | Benefício | Peso |
|---------------------------------|--|--|--------------|
| | Fornecer indicador de versão (versão legível por computador) |   | 40 |
| | Fornecer o histórico de versão (versão legível por humano) |   | 40 |
| | Fornecer o histórico de versão (versão legível por computador) |   | 40 |
| Identificadores de Dados | Usar URIs persistentes como identificadores de conjuntos de dados |    | 60 |
| | Usar URIs persistentes como identificadores dentro de conjuntos de dados |     | 80 |
| | Atribuir URIs para as versões dos conjuntos de dados e séries |    | 60 |
| Formato de Dados | Usar formatos de dados padronizados e legíveis por máquinas |   | 40 |
| | Usar representações de dados que sejam independentes de localidade (<i>locale neutral</i>) (versão legível por humano) |   | 40 |
| | Usar representações de dados que sejam independentes de localidade (<i>locale neutral</i>) (versão legível por computador) |   | 40 |
| | Fornecer dados em vários formatos |   | 40 |
| Vocabulário de Dados | Reutilizar vocabulários, dando preferência aos padronizados |      | 100 |
| | Escolher o nível de formalização adequado |    | 60 |
| Acesso aos Dados | Fornecer download em massa (<i>bulk download</i>) |   | 40 |
| | Usar negociação de conteúdo para servir os dados disponíveis em vários formatos |   | 40 |
| | Fornecer dados atualizados |   | 40 |
| | Fornecer uma explicação para os dados que não estão disponíveis (versão legível por humano) |   | 40 |
| | Fornecer uma explicação para os dados que não estão disponíveis (versão legível por computador) |   | 40 |
| Feedback | Recolher <i>feedback</i> dos consumidores de dados |    | 60 |
| | Compartilhar o <i>feedback</i> disponível |   | 40 |
| Enriquecimento de Dados | Fornecer visualizações complementares |     | 80 |
| Replicação de Dados | Obedecer os termos de licença |   | 40 |
| | Citar a publicação original do conjunto de dados |    | 60 |
| TOTAL | | | 1500 100% |

Nota: *R= Reuso; C= Compreensão; L= Capacidade de conexão; D= Facilidade de descoberta; T= Confiança; A= Acesso; I= Interoperabilidade; P= Processabilidade

A partir desta pontuação, conclui-se que a máxima pontuação que um conjunto de dados consegue obter é 1500 pontos (100%), pois terá de contemplar todas as boas práticas.

Para classificar cada conjunto de dados nas classificações de mau, suficiente, bom e excelente, utilizou-se o seguinte critério de acordo com a percentagem de pontos alcançada por um conjunto de dados:

- a) classificação de mau varia entre 0% a 25%;
- b) classificação de suficiente varia entre 26% e 50%;
- c) classificação de bom varia entre 51% e 75%; e
- d) classificação de excelente varia entre 76% e 100%;

Para a classificação percentual e qualitativa de cada conjunto de dados, foi utilizado uma folha Excel, onde para cada conjunto de dados se apresenta a pontuação final, o percentual de pontos e a sua classificação qualitativa. Para cada BP atendida foi associado o número um, em caso negativo ou não aplicado, foi associado zero.

4. RESULTADOS E DISCUSSÃO

Os resultados são apresentados em dois momentos, a) correspondência entre os indicadores analisados pelo EDP e as DWBP e a correspondência entre as DWBP e os indicadores do EDP, e b) análise de correspondência. Na análise de correspondência entre os indicadores do EDP e as DWBP, encontra-se as dimensões analisadas pelo EDP e dentro de cada dimensão estão os indicadores que são avaliados nas mesmas. Para cada dimensão é indicada a/as boas práticas que lhes corresponde totalmente. Na análise de correspondência entre as DWBP e os indicadores do EDP é apresentado para cada boa prática, os indicadores avaliados pelo EDP que lhes corresponde.

A fase de análise de correspondência apresenta o tipo de correspondência que existe entre os indicadores do EDP e as DWBP e vice-versa.

Na correspondência entre as DWBP e os indicadores do EDP, as boas práticas que tiveram análise de correspondência total com algum indicador já avaliado pelo EDP serão alvo de confronto com a análise da qualidade dos dados do EDP realizada à luz das DWBP.

4.1 Correspondência entre Indicadores EDP e Indicadores DWBP

Nesta seção é apresentada a análise de correspondência entre os indicadores do EDP e as DWBP e a análise de correspondência entre as DWBP e os indicadores do EDP, bem como o tipo de correspondência adotado. As correspondências entre os indicadores do EDP com as DWBP aparecem marcadas a azul-escuro. As correspondências entre as DWBP com os indicadores do EDP encontram-se marcados a azul-claro e as correspondências em comum entre as duas análises aparecem marcadas a verde.

As análises de correspondência entre indicadores do EDP e as DWBP e vice-versa, são apresentadas no Quadro 1.

Quadro 1 - Correspondência entre indicadores EDP e indicadores DWBP

| | BP1 | BP2 | BP3 | BP4 | BP5 | BP6 | BP7 | BP8 | BP9 | BP10 | BP11 | BP12 | BP13 | BP14 | BP15 | BP16 | BP17 | BP18 | BP19 | BP20 | BP21 | BP22 | BP23 | BP24 | BP25 | BP26 | BP27 | BP28 | BP29 | BP30 | BP31 | BP32 | BP33 | BP34 | BP35 | | |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|--|
| Pesquisabilidade | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Utilização de palavra-chave | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Categorias | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pesquisa Espacial | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pesquisa Temporal | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Acessibilidade | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Acessibilidade da URL de acesso | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| URL de transferência | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Acessibilidade da URL de transferência | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Interoperabilidade | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Formato | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tipo de Suporte | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Formato/tipo de suporte do vocabulário | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Não proprietário | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Legível por máquina | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Conformidade com o sistema DCAT AP | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Reusabilidade | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Informações constantes da licença | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Vocabulário da licença | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Restrições de acesso | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Vocabulário das restrições de acesso | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Ponto de contacto | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Publicador | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Contextualidade | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Direitos | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tamanho do ficheiro | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data de emissão | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data de alteração | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Legenda: ■ Correspondência total entre os indicadores do EDP com as DWBP; ■ Correspondência total entre as DWBP e os indicadores do EDP; ■ Correspondência comum entre as duas análises

Após comparar as correspondências entre os indicadores do EDP e as DWBP, conclui-se que, ao corresponder os indicadores do EDP com as DWBP encontram-se as seguintes análises de correspondência totais:

- a) o indicador utilização de palavra-chave mapeia totalmente com a BP2- fornecer metadados descritivos; a pesquisa espacial mapeia totalmente com a BP1- fornecer metadados;
- b) a pesquisa temporal mapeia com a BP5- fornecer informações de proveniência de dados e com a BP21- fornecer dados atualizados;
- c) a acessibilidade da URL de acesso mapeia com a BP9- use URIs persistentes como identificadores de conjuntos de dados e com a BP10- use URIs persistentes como identificadores em conjuntos de dados;
- d) a URL de transferência mapeia com a BP9 e com a BP10;
- e) a acessibilidade da URL de transferência mapeia totalmente com a BP9 e com a BP10;
- f) o formato mapeia com a BP14- fornecer dados em vários formatos;
- g) o tipo de suporte mapeia com a BP23- disponibilizar dados através de uma API;
- h) o formato/tipo de suporte de vocabulário mapeia com a BP15- reutilize vocabulários de preferência padronizados e com a BP16- escolha o nível de formalização correto;
- i) o indicador legível por máquina mapeia com a BP12- use formatos legíveis por máquina;
- j) o indicador informações constantes da licença mapeia com a BP4- fornecer informações de licença de dados;
- k) o vocabulário da licença mapeia com a BP15; o ponto de contacto mapeia totalmente com a BP29-reunir *feedback* dos consumidores de dados;
- l) o publicador mapeia com a BP33- fornecer *feedback* ao editor original;
- m) o tamanho do ficheiro mapeia com a BP1;
- n) a data de emissão mapeia com a BP1; e
- o) a data de alteração mapeia com a BP1.

No que diz respeito à correspondência entre as DWBP com os indicadores analisados pelo EDP, encontram-se as seguintes análises de correspondência totais:

- a) a BP1 mapeia com a utilização de palavra-chave;
- b) a BP2 mapeia com o indicador categorias;

- c) a BP4 é coberta pelo indicador informações constantes da licença;
- d) a BP9 mapeia totalmente com a acessibilidade da URL de acesso, a URL de transferência e a Acessibilidade da URL de transferência;
- e) a BP10 mapeia com a acessibilidade da URL de acesso, a URL de transferência e a Acessibilidade da URL de transferência;
- f) a BP11 mapeia com a acessibilidade da URL de acesso, a URL de transferência e a Acessibilidade da URL de transferência;
- g) a BP12 mapeia com o indicador legível por máquina;
- h) a BP14 mapeia com o formato; a BP15 mapeia com o formato/tipo de suporte de vocabulário;
- i) a BP21 faz correspondência com a data de emissão e com a data de alteração; e
- j) a BP29 mapeia totalmente com o ponto de contacto.

Na análise de correspondência dos indicadores do EDP com as DWBP observa-se que existem 17 indicadores avaliados pelo EDP que são mapeados totalmente pelas boas práticas para a Web (DWBP).

Relativamente à correspondência entre as DWBP e os indicadores do EDP observa-se que existem 11 boas práticas que são mapeadas totalmente por um indicador do EDP. Além das 11 boas práticas mapeadas pelo EDP, este estudo avançou na análise das outras 14 práticas.

Na correspondência entre indicadores do EDP e DWBP e correspondência entre DWBP e indicadores do EDP, observa-se ainda que existem análises de correspondência totais que coincidem com as duas análises. As análises de correspondência em comum entre as duas análises são: a BP4 que mapeia com as informações constantes da licença; a BP9 e BP10 que mapeiam com a acessibilidade da URL de acesso, com a URL de transferência e com a acessibilidade da URL de transferência; a BP12 que mapeia com o indicador legível por máquina; a BP14 que mapeia com o formato; a BP15 que mapeia com o formato/tipo de suporte de vocabulário; e a BP29 que mapeia com o ponto de contacto.

4.2 Análise do Portal: Estudo Exploratório

O estudo exploratório incidiu sobre os 20 primeiros conjuntos de dados nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto. Assim, ordenados

pele último modificado e filtrados pela classificação de Excelente ou Bom+ verificou-se se os conjuntos de dados analisados contemplavam as 35 boas práticas para a Web (Anexo 1).

Na categoria de Governo e Setor Público, com a classificação de excelente, as BP1, BP2, BP4, BP5, BP6, BP7, BP8, BP10, BP11, BP12, BP14, BP15, BP16, BP19, BP21, BP29, BP34 e a BP35 foram detetadas em todos os conjuntos de dados; a BP3 e BP22 não foram identificadas em nenhum conjunto de dados; a BP9, BP13, BP30 e a BP32 não se apresentaram em nenhum conjunto de dados analisados. A BP17 está contemplada por 12 conjuntos de dados e 8 conjuntos não adotaram.

Na Figura 15 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são adotadas.

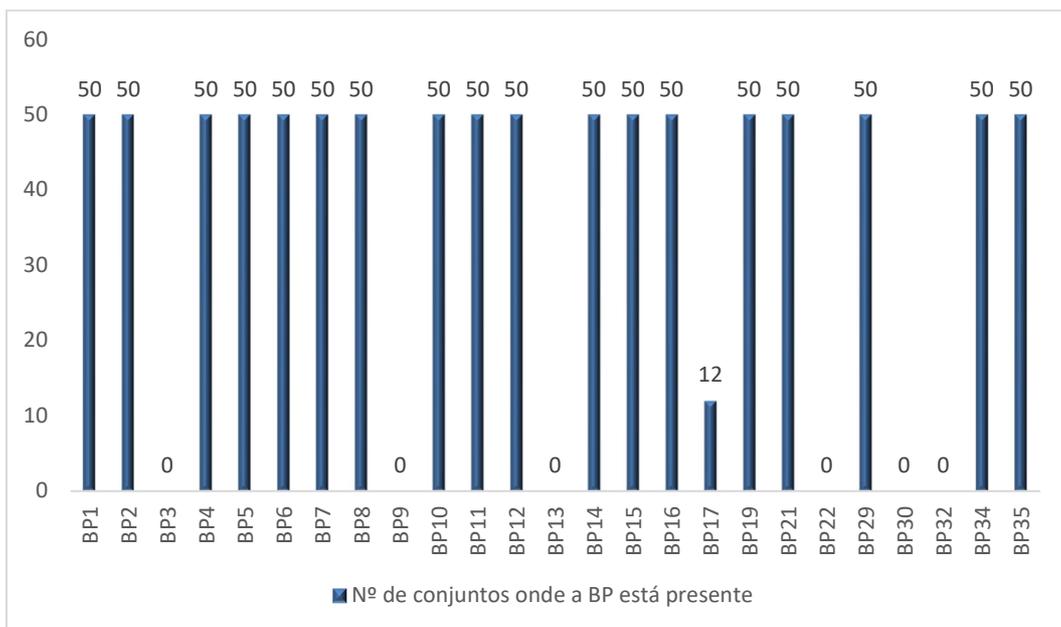


Figura 15 - Número de conjuntos de dados da categoria de Governo e Setor Público em que cada BP está presente.

Na categoria Transportes, com a classificação de Bom+, as BP1, BP2, BP4, BP5, BP6, BP11, BP12, BP21 e BP34 foram contempladas por todos os conjuntos de dados; as BP3, BP16 e BP22 não foram identificadas em nenhum conjunto analisado; a BP7 foi observada em 2 conjuntos de dados analisados e em 18 não; as BP8, BP13, BP30 e BP32 não foram observadas em nenhum conjunto de dados; a BP9 foi identificada em 5 conjuntos e não identificada em 15 conjuntos de dados.

A BP10 não foi identificada em 19 conjuntos de dados e não foi detetada em 1; em 13 conjuntos de dados não contemplaram a BP14 e 7 contemplavam; a BP15 não foi observada em nenhum dos conjuntos de dados; a BP17 foi encontrada em 5 conjuntos de dados e 15 não; para a BP19 foram encontrados 8 conjuntos onde a boa prática não foi identificada, 10 conjuntos que não contemplavam a BP e 2 que adotaram a BP; a BP29 estava contemplada em 7 conjuntos de dados e não contemplada em 13; na BP35 foram identificados 17 conjuntos de dados onde esta boa prática não estava presente e 3 conjuntos onde esta estava presente.

Na Figura 16 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são adotadas.

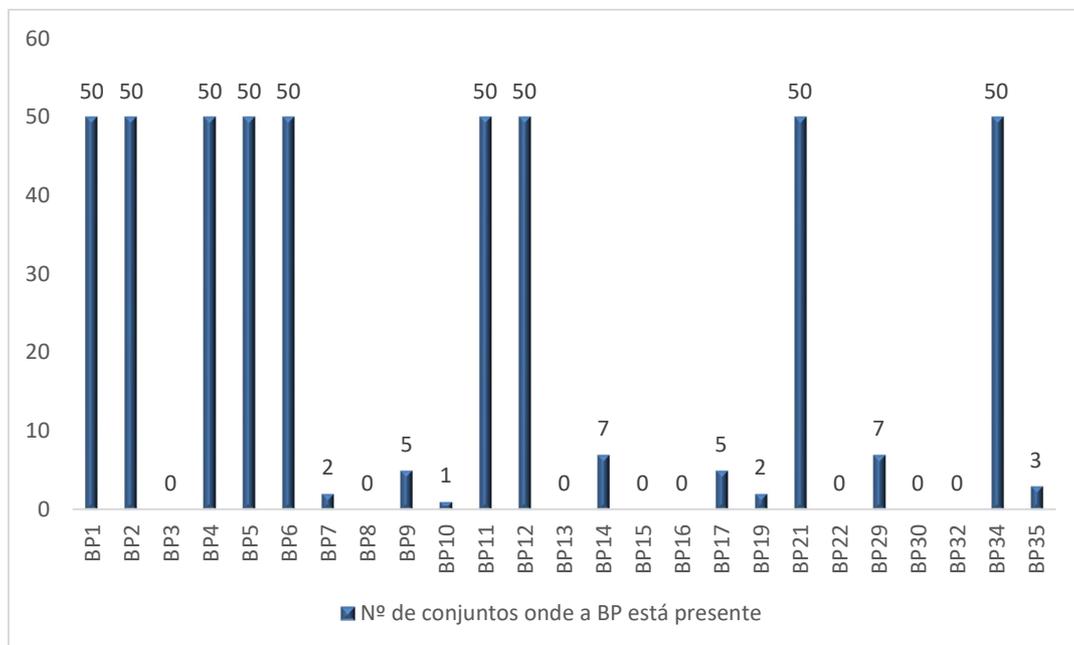


Figura 16 - Número de conjuntos de dados da categoria de Transportes em que cada BP está presente.

Na categoria Educação, Cultura e Desporto, com a classificação de Excelente e Bom+, a BP1, BP2, BP4, BP5, BP6, BP11, BP12, BP21 e BP34 identificaram-se em todos os conjuntos de dados; a BP3, BP10, BP16 e B22 não foram identificadas em nenhum conjuntos de dados; a BP7 verificou-se em 2 conjuntos de dados e não se verificou em 18; a BP8, BP13, BP19, BP30, BP32 e BP35 não estavam presentes em nenhum conjuntos de dados; a BP9 estava presente em 15 conjuntos de dados e não estava presente em 5; na BP14 foram encontrados 19 conjuntos de dados que observavam a boa prática e 1 que não; na BP15 foram encontrados

19 conjuntos de dados que não contemplavam a boa prática e 1 que contemplava; na BP17, 8 conjuntos de dados analisados não contemplavam a boa prática e 12 contemplavam; a BP29 não foi contemplada em 7 dos conjuntos analisados e foi identificada em 13.

Na Figura 17 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são aplicadas.

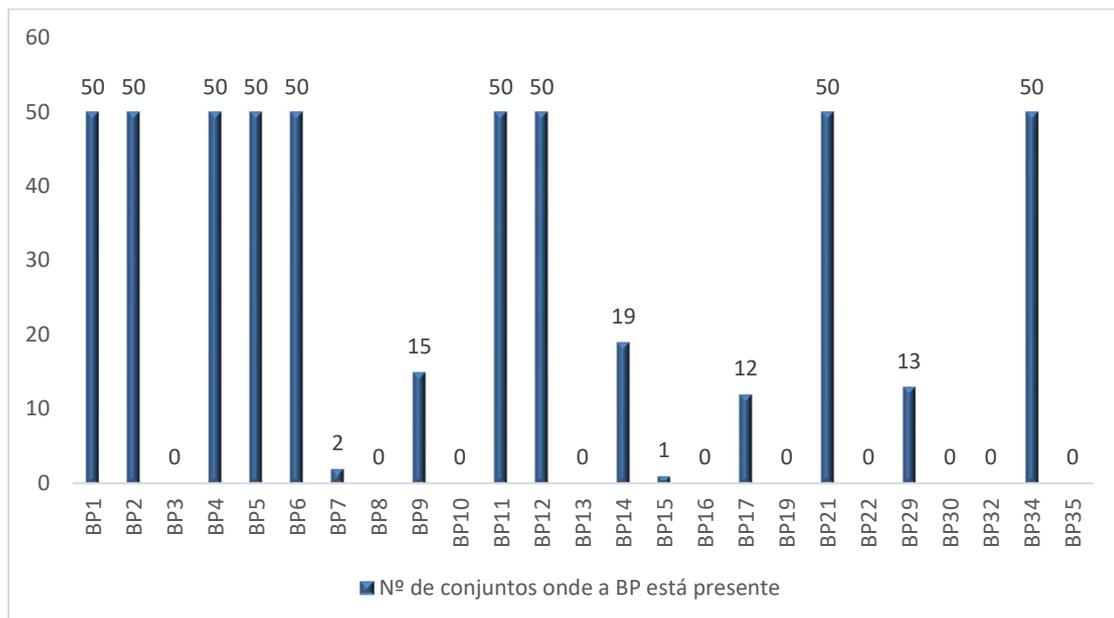


Figura 17 - Número de conjuntos de dados da categoria de Educação Cultural e Desporto em que cada BP está presente.

Com a realização da análise de correspondência entre as boas práticas para a Web e os indicadores do EDP, conclui-se que algumas das boas práticas já são cobertas por algum indicador de uma das dimensões do EDP.

Assim, do conjunto das DWBP, existem 11 boas práticas que se encontram correspondidas totalmente com um indicador do EDP, sendo elas as seguintes: BP1, BP2, BP4, BP9, BP10, BP11, BP12, BP14, BP15, BP21 e BP29.

A realização deste estudo exploratório permitiu confrontar se os resultados obtidos com a análise dos conjuntos de dados do EDP à luz das DWBP coincidem com os resultados obtidos pelo próprio EDP com base nas boas práticas mapeadas totalmente por um indicador próprio.

Assim, relativamente à BP1, fornecer metadados (esta boa prática é coberta pelo EDP e de todos os conjuntos de dados analisados de todas as categorias), todos cumpriam com a

boa prática. O mesmo acontece com a BP2, fornecer meadados descritivos e a BP4, fornecer informações sobre licença dos dados, BP11 atribuir URIs a versões e séries de conjuntos de dados.

No que diz respeito às BP9, BP10 e BP11, estas são mapeadas totalmente pelo EDP, no entanto, na boa prática 9. Na BP10, na maioria dos conjuntos de dados não foi possível encontrar URIs dentro dos conjuntos de dados, pelo que não foi possível identificar se esta boa prática está a ser cumprida.

Todos os conjuntos de dados analisados, nas três categorias, cumpriam com a BP12, fornecer formatos legíveis por máquina, também coberta pelo EDP.

A boa prática 14, fornecer dados em vários formatos é mapeada totalmente pelo EDP, no entanto surgiram 14 conjuntos de dados que não apresentavam os dados fornecidos em vários formatos, contendo apenas uma ou duas distribuições.

A boa prática 15, reutilizar vocabulários preferencialmente padronizados, é coberta pelo EDP, porém, mais de metade dos conjuntos de dados analisados não reutiliza vocabulários.

No que diz respeito à boa prática 21, fornecer dados atualizados, esta está coberta pelo EDP e todos os conjuntos de dados possuem dados atualizados.

Por fim, a boa prática 29, reunir *feedback* dos consumidores de dados, está coberta pelo EDP, contudo, 20 conjuntos de dados não possuem nenhum ponto de contacto.

Em suma, com estes resultados averiguou-se que as boas práticas 1, 2, 4, 11, 12 e 21 são iguais aos resultados apresentados pelo European Data Portal, no entanto, nas boas práticas 9, 10, 11, 14, 15 e 29 os resultados estão relativamente perto dos resultados do EDP, mas ainda assim, existem alguns conjuntos de dados onde os resultados divergem do EDP.

4.3 Análise do Portal: estudo final

A análise do estudo final foi realizada nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Governo. Assim, ordenado pelo último conjunto de dados modificado e filtrados por Excelente ou Bom+, foram analisados os 50 primeiros conjuntos de cada categoria de acordo com a técnica de amostragem referida anteriormente, de maneira a verificar se estes contemplavam as 35 boas práticas do W3C.

Para a categoria de Governo e Setor Público, com a classificação de Bom+:

- a) a BP1, BP4, BP17, BP19 e BP34 foram observadas em todos os conjuntos analisados;
- b) a BP2 foi observada por 38 conjuntos de dados e não por 12;
- c) a BP3 não foi observada em nenhum conjunto de dados;
- d) a BP5 foi observada por 49 conjuntos de dados e não por 1;
- e) a BP6 não foi observada em nenhum conjunto de dados;
- f) a BP7H foi observada por 6 conjuntos de dados e não por 44;
- g) a BP7C foi observada por 5 conjuntos de dados e não por 45;
- h) a BP8H foi observada por 40 conjuntos de dados e não por 10;
- i) a BP8C foi observada por 5 conjuntos de dados e não por 45;
- j) na BP9 não foi observada em nenhum conjunto de dados;
- k) na BP10 foi observada por 48 conjuntos de dados e não por 2;
- l) a BP11 foi observada por 49 conjuntos de dados e não por 1;
- m) a BP12 foi observada por 49 conjuntos de dados e não por 1;
- n) a BP13H foi observada por 17 conjuntos de dados e não por 33;
- o) a BP13C foi observada por 37 conjuntos de dados e não por 13;
- p) a BP14 foi observada por 19 conjuntos de dados e não por 31;
- q) a BP15 foi observada por 18 conjuntos de dados e não por 32;
- r) a BP16 foi observada por 49 conjuntos de dados e não por 1;
- s) a BP21 foi observada por 23 conjuntos de dados e não por 27;
- t) a BP22H não foi observada em nenhum conjunto de dados;
- u) a BP22C não foi observada em nenhum conjunto de dados;
- v) a BP29 foi observada por 42 conjuntos de dados e não por 8;
- w) a BP30 não foi observada em nenhum conjunto de dados;
- x) a BP32 foi observada por 23 conjuntos de dados e não por 27; e
- y) a BP35 foi observada por 7 conjuntos de dados e não por 43 (Anexo 2).

Na Figura 18 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são aplicadas.

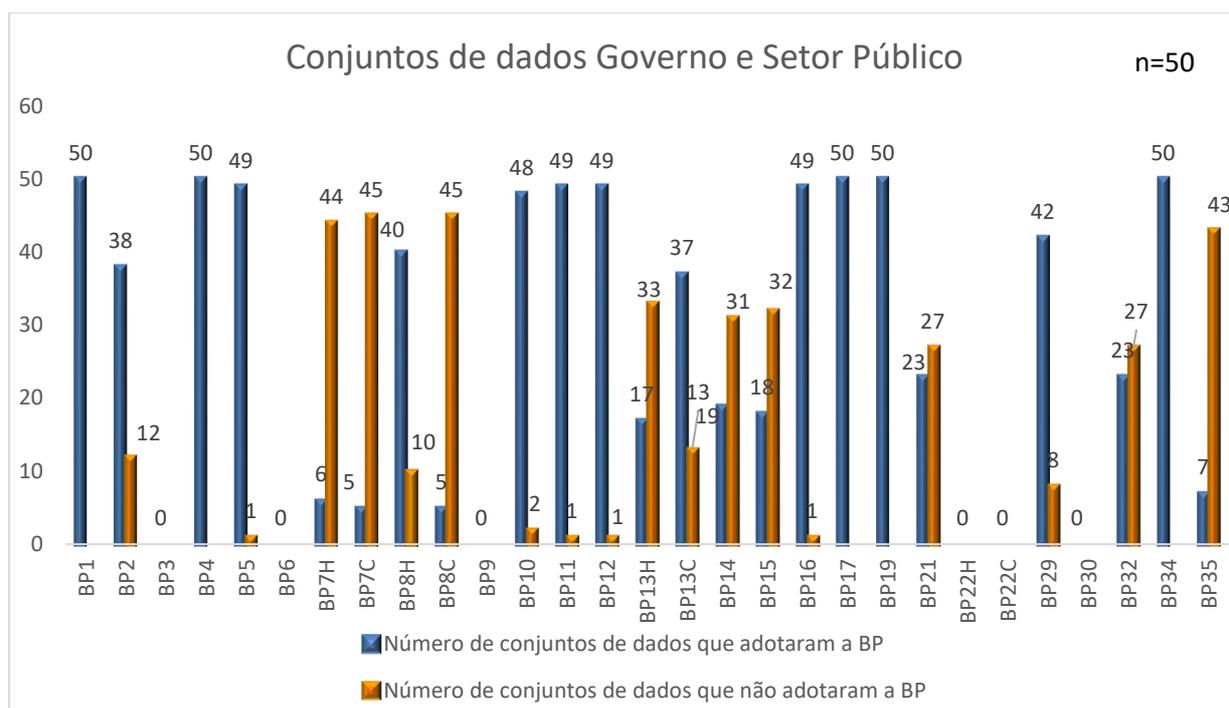


Figura 18 - Número de conjuntos de dados da categoria de Governo e Setor Público em que cada BP está presente.

Na categoria de Transportes, com a classificação de Bom+:

- a) as BP1, BP4, BP5, BP10, BP11, BP12, BP16, BP17, BP19 e BP34 foram observadas em todos os conjuntos de dados analisados;
- b) a BP2 foi observada por 45 conjuntos de dados e não por 5;
- c) a BP3 foi observada por 9 conjuntos de dados e não por 41;
- d) a BP6 não foi observada por nenhum conjunto de dados;
- e) a BP7H foi observada por 16 conjuntos de dados e não por 34;
- f) a BP7C foi observada por 16 conjuntos de dados e não por 34;
- g) a BP8H foi observada por 45 conjuntos de dados e não por 5;
- h) a BP8C não foi observada em nenhum conjunto de dados;
- i) a BP9 não foi observada em nenhum conjunto de dados;
- j) a BP13H foi observada por 9 conjuntos de dados e não por 41;
- k) a BP13C foi observada por 36 conjuntos de dados e não por 14;
- l) a BP14 foi observada por 21 conjuntos de dados e não por 29;
- m) a BP15 foi observada por 2 conjuntos de dados e não por 48;
- n) a BP21 foi observada por 10 conjuntos de dados e não por 40;
- o) a BP22H não foi observada em nenhum conjunto de dados;

- p) a BP22C não foi observada em nenhum conjunto de dados;
- q) a BP29 foi observada por 44 conjuntos de dados e não por 6;
- r) a BP30 não foi observada em nenhum dos conjuntos de dados;
- s) a BP32 foi observada por 35 conjuntos de dados e não por 17; e
- t) a BP35 foi observada por 4 conjuntos de dados e não por 46 (Anexo 3).

Na Figura 19 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são aplicadas.

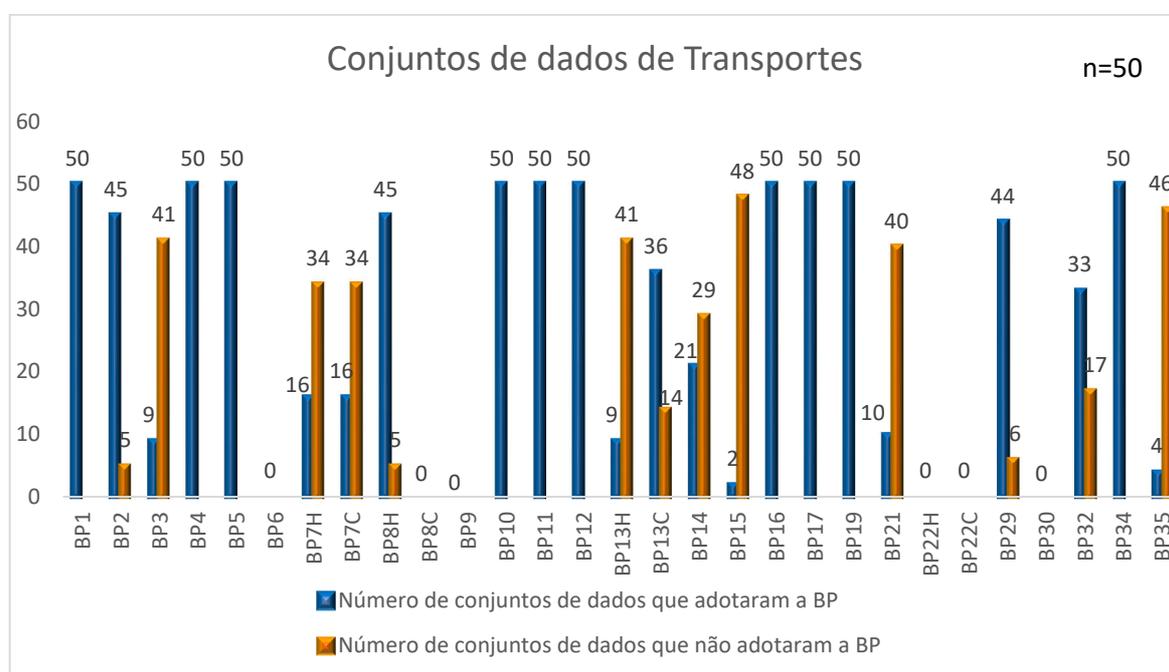


Figura 19 - Número de conjuntos de dados da categoria de Transporte em que cada BP está presente.

Na categoria de Educação, Cultura e Desporto, com a classificação de Bom+:

- a) as BP1, BP4, BP5, BP11, BP16, BP17, BP19 e BP34 observaram-se em todos os conjuntos de dados;
- b) a BP2 foi observada em 46 conjuntos de dados e não em 4;
- c) a BP3 foi observada em 4 conjuntos de dados e não em 46;
- d) a BP6 não foi observada por nenhum conjunto de dados;
- e) a BP7H foi observada em 5 conjuntos de dados e não em 45;
- f) a BP7C foi observada em 5 conjuntos de dados e não em 45;

- g) a BP8H foi observada em 47 conjuntos de dados e não em 3;
- h) a BP8C foi observada em 5 conjuntos de dados e não em 45;
- i) a BP9 não foi observada em nenhum conjunto de dados;
- j) a BP10 foi observada em 49 conjuntos de dados e não em 1;
- k) a BP12 foi observada em 48 conjuntos de dados e não em 2;
- l) a BP13H foi observada em 14 conjuntos de dados e não em 36;
- m) a BP13 foi observada em 31 conjuntos de dados e não em 19;
- n) a BP14 foi observada em 22 conjuntos de dados e não em 28;
- o) a BP15 foi observada em 1 conjunto de dados e não em 49;
- p) a BP21 foi observada em 17 conjuntos de dados e não em 33;
- q) a BP22H não foi observada por nenhum conjunto de dados;
- r) a BP22C não foi observada por nenhum conjunto de dados;
- s) a BP29 foi observada em 33 conjuntos de dados e não em 17;
- t) a BP30 não foi observada por nenhum dos conjuntos de dados;
- u) a BP32 foi observada em 24 conjuntos de dados e não em 26; e
- v) a BP35 foi observada em 7 conjuntos de dados e não em 43 (Anexo 4).

Na Figura 20 é possível observar para esta categoria, quais as BP's contempladas com maior e menor frequência pelos conjuntos de dados. A Figura não mostra as BP's que não são aplicadas.

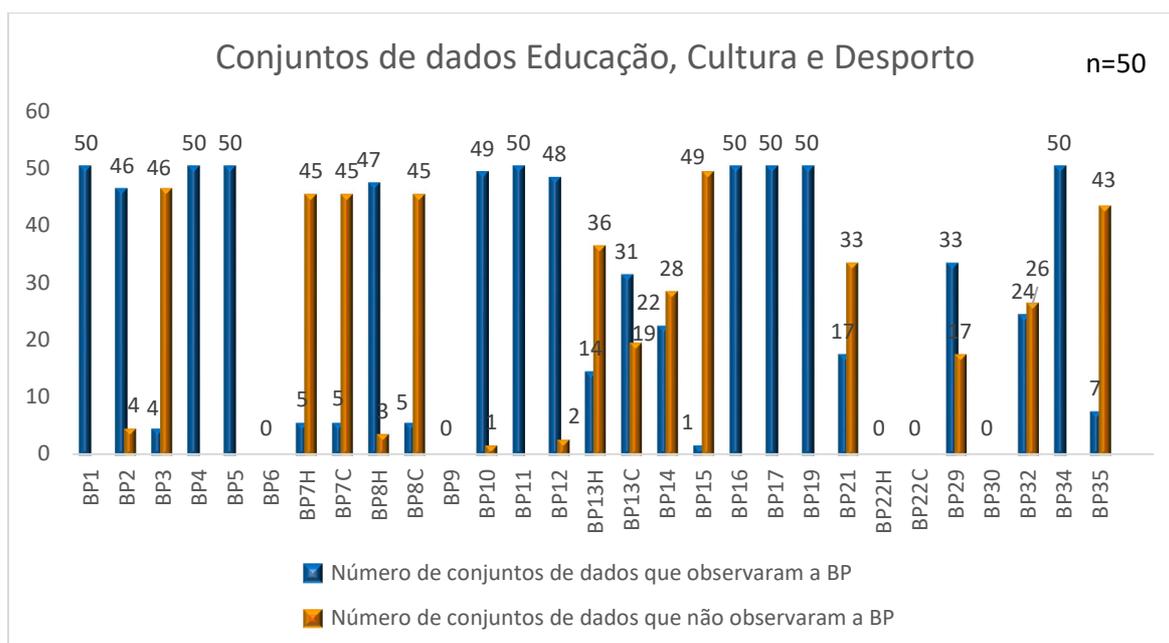


Figura 20- Número de conjuntos de dados da categoria de Educação, Cultura e Desporto em que cada BP está presente.

Segundo Simperl & Walker (2017), para que um portal de dados atinga a sustentabilidade e valor acrescentado este tem de cumprir com 10 requisitos, entre eles os seguintes: ser mensurável e publicar metadados. Quando comparado com as DWBP, estas também verificam que o conjunto de dados apresenta metadados e informações de qualidade dos dados. Na análise destas BP's verificou-se que estas eram cumpridas por todos os conjuntos de dados, uma vez que, o EDP já obrigava ao fornecimento de metadados e já disponibilizava os conjuntos de dados distribuídos pela qualidade dos mesmos, o que vem apoiar a afirmação do autor.

Como referido anteriormente, algumas das DWBP são cobertas por indicadores do EDP, sendo elas as seguintes: as BP1, BP2, BP4, BP9, BP10, BP11, BP12, BP14, BP15, BP21 e BP29.

Para realizar o confronto entre as boas práticas mapeadas com os resultados do EDP, foi realizada a soma de todos os conjuntos de dados analisados das três categorias (150), verificando quantos adotaram ou não adotaram a boa prática mapeada, de forma a alcançar uma visão mais geral.

Em relação às BP1 e BP4 que já são mapeadas pelo EDP todos os conjuntos de dados analisados das três categorias tinham as boas práticas presentes, o que significa que os resultados estão de acordo com o EDP. A BP11 estava presente em todos os conjuntos de dados, com a exceção de um na categoria de Governo e Setor Público.

Tal como referido por Huyer & Cecconi (2018), para a reutilização dos dados é necessário ter em consideração as condições em que estas reutilizações podem ocorrer, a exemplo da sua publicação sob uma licença aberta. Mas que isso, a identificação do tipo de licença possibilita o reuso com maior segurança (Lóscio et al., 2017, Torino et al., 2019).

As DWBP também preveem que os conjuntos de dados possuam licença, e de todos os conjuntos de dados analisados, todos eles possuíam licença. No entanto, em certos conjuntos não era especificado o tipo de licença atribuída o que pode comprometer a forma como pode ser reutilizado.

Para a BP2, que também é contemplada pelo EDP, surgiram casos onde esta não estava a ser cumprida, tendo sido encontrados 129 conjuntos de dados que atendiam a boa prática e 21 que não atendiam.

Para a BP9, não foram encontrados conjuntos onde esta BP se observou.

Na BP10, surgiram 147 conjuntos de dados onde esta BP foi observada e 3 que não adotavam.

Para a BP12, surgiram 147 conjuntos de dados onde esta BP foi observada e 3 que não adotavam.

Relativamente à BP14, que também é coberta pelo EDP, foram encontrados 88 conjuntos de dados que não observaram esta BP e 62 que observaram.

Para a BP15 foram encontrados 21 conjuntos de dados que adotaram a boa prática e 129 que não.

Para a BP21 foram encontrados 50 conjuntos de dados que adotaram a boa prática e 100, não.

Por último, para a BP29, foram encontrados 119 conjuntos de dados que cumpriam com a boa prática e 31 que não cumpriam.

Assim, das DWBP mapeadas apenas as BP1, BP4 e BP11 estão de acordo com os resultados chegados pelo EDP.

As BP2, BP10, BP12 e BP29 estão perto do resultado chegado pelo EDP com a maioria dos resultados atingidos iguais aos do EDP e apenas alguns que não estão de acordo.

A BP 29 refere-se a coleta de *feedback*, e como referido por Knippenberg (2015), existem cada vez mais portais de dados que permitem a interação entre editores e reutilizadores de dados através do *feedback* dos dados. As DWBP apresentam realmente uma melhor prática que verifica se os conjuntos de dados apresentam um email ou outra forma de contacto, que

permita reunir *feedback* dos consumidores de dados. Contudo, foram encontrados alguns conjuntos de dados que não apresentavam nenhuma forma que permitisse aos consumidores darem *feedback* dos dados, o que vem contrariar a afirmação de Knippenberg (2015).

Já nas BP9, BP14, BP15 e BP21 foram mais os conjuntos de dados encontrados que não se alinham a estas boas práticas ou onde estas não foram observadas, do que os casos onde estas foram encontradas.

Em forma de conclusão, das 11 DWBP mapeadas pelo EDP apenas 3 (BP1, BP4 e BP11) estão de acordo com os resultados do EDP, 4 (BP2, BP10, BP12 e BP29) estão perto dos resultados atingidos, mas ainda assim com algumas discrepâncias, enquanto nas outras 4 (BP9, BP14, BP15 e a BP21) foram encontrados resultados que representam quase o oposto do resultado atingido pelo EDP.

O estudo do portal permitiu fazer uma classificação em relação aos conjuntos de dados, como explicitado no tópico 3.1.

No estudo final foram analisadas as seguintes dimensões: a dimensão dos metadados, a licença de dados, a proveniência e qualidade; o versionamento de dados, os identificadores de dados, o formato de dados, o vocabulário de dados, o acesso aos dados, a preservação de dados, o *feedback*, o enriquecimento de dados e a republicação de dados. Dentro de cada dimensão, são analisadas as DWBP, como referido na Tabela 1.

Como referido anteriormente, a atribuição de uma pontuação para os conjuntos de dados analisados foi vinculada ao benefícios de cada BP. De acordo com a Tabela 1, cada BP traz certos benefícios, e a pontuação de cada BP difere de acordo com os benefícios da mesma.

Como não foram encontrados estudos científicos, até a finalização deste estudo, que comprovassem que um benefício era melhor que o outro, ou que uma BP deveria ter maior pontuação porque tem mais benefícios, por exemplo, decidiu-se que cada benefício iria receber a mesma pontuação. Assim, cada BP recebe a pontuação de acordo com o número de benefícios obtidos.

A soma de todas as BP's presentes num determinado conjunto de dados, representa a sua pontuação final. A pontuação máxima que um conjunto de dados pode atingir são 1500 pontos, uma vez que existem BP's que apresentam versões legíveis por humano e por máquina, o que duplica a pontuação da BP e não foram contabilizadas as BP's que não eram aplicadas, de forma a não interferir com a classificação final.

A classificação de cada conjunto de dados depende da pontuação final que este venha a atingir. A classificação final tem lugar por meio de quatro grupos de classificação. O mapeamento dos pontos para a categoria de classificação é indicado na Tabela 2.

Tabela 3 - Escala percentual para cada classificação

| Classificação | Percentual de pontos atingido |
|----------------------|--------------------------------------|
| Excelente | 76% a 100% |
| Bom | 51% a 75% |
| Suficiente | 26% a 50% |
| Mau | 0% a 25% |

Para a categoria de Governo e Setor Público, os conjuntos de dados analisados segundo as DWBP apresentam a classificação descrita na Tabela 3.

Tabela 4 - Classificação dos conjuntos de dados na categoria de Governo e Setor Público

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|------------------------|---------------------------------------|----------------------|
| https://data.europa.eu/data/datasets/eu-whoiswho-the-official-directory-of-the-european-union?locale=pt | 1000 | 67% | Bom |
| http://data.europa.eu/88u/dataset/covid-19-multilingual-terminology-on-iate | 760 | 51% | Bom |
| http://data.europa.eu/88u/dataset/efgs | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/official-journals-of-the-european-union-2011_14 | 560 | 37% | Suficiente |
| https://data.europa.eu/data/datasets/https-datosabiertos-ayto-arganda-es-dataset-actas-de-pleno-municipal-legislatura-2019-2023 | 800 | 53% | Bom |
| http://data.europa.eu/88u/dataset/public-opinion-in-the-eu-in-time-of-coronavirus-crisis-2 | 680 | 45% | Suficiente |
| https://data.europa.eu/data/datasets/eu-open-data-portal-all-datasets-and-their-metadata | 640 | 43% | Suficiente |
| http://data.europa.eu/88u/dataset/5X5vtt7NouitY Ys514vTA | 700 | 47% | Suficiente |
| https://data.europa.eu/data/datasets/h1eq482pt5 y647mvdndsg | 700 | 47% | Suficiente |
| http://data.europa.eu/88u/dataset/QE0oTUz7zFW eAzy6d2Lg | 700 | 47% | Suficiente |
| http://data.europa.eu/88u/dataset/YsRglGg1ve4y ELUuWuI9bA | 700 | 47% | Suficiente |
| http://data.europa.eu/88u/dataset/tjQavm4mN6Zl yvasfitQ5Q | 700 | 47% | Suficiente |
| http://data.europa.eu/88u/dataset/1OEJUOBj20A eH2oP5GFObw | 700 | 47% | Suficiente |
| http://data.europa.eu/88u/dataset/vZelr8PWRdG njpiyhEOXw | 700 | 47% | Suficiente |

Tabela 5 - Classificação dos conjuntos de dados na categoria de Governo e Setor Público (Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| http://data.europa.eu/88u/dataset/class-sum-leg | 1000 | 67% | Bom |
| http://data.europa.eu/88u/dataset/file-status | 1000 | 67% | Bom |
| http://data.europa.eu/88u/dataset/legal-act-domain | 940 | 63% | Bom |
| http://data.europa.eu/88u/dataset/role-qualifier | 1000 | 67% | Bom |
| http://data.europa.eu/88u/dataset/european-commission-hr-key-figures-2018 | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-81168-de?locale=fi | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen236-c-8202-de?locale=fr | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-62365-de?locale=hr | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-63348-de?locale=lt | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-37055-de | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-37029-de/?locale=pt | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-62634-de?locale=en | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen53-c-https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen236-c-8249-de?locale=ga | 960 | 64% | Bom |
| https://data.europa.eu/data/datasets/a7dcca82-beda-4db2-9767-59dc42d9ac3a | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/845089c0-2ec3-4ce4-b074-118826d8e1d1 | 940 | 63% | Bom |
| https://data.europa.eu/data/datasets/b1ef448f-6016-4a08-a54a-3bebebd487c9 | 880 | 59% | Bom |
| https://data.europa.eu/data/datasets/c55386e0-5bc0-4b89-9a07-62d57a194a0e | 880 | 59% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-30968f0dd0c0499a849d8a274837f4f7_0?locale=pt | 940 | 63% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-b0903f204e4e46e598ce971657a45472_0?locale=pt | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-b0903f204e4e46e598ce971657a45472_0?locale=pt | 820 | 55% | Bom |

Tabela 6 - Classificação dos conjuntos de dados na categoria de Governo e Setor Público (Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-86ec319aeeef842048cd3260ae01bd3f9_0?locale=pt | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-fabdef0019fc4dc5ae8e834bbc9491b4_0?locale=pt | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-1a25a44228754bd1b2c8de4b995e8e87_0?locale=pt | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/stanord_cm_s-61037?locale=pt | 880 | 59% | Bom |
| https://data.europa.eu/data/datasets/021ce124-2ac0-4aee-a3d8-a1d33932b2ac | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/8c953125-eec4-48c6-a562-416d36e7920f | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/36e3596b-71d4-4bcb-bf03-040e662b5136 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/227b95a4-f1f0-4188-b724-acb3eba36069 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/ccaafa29-7055-4d14-933a-2f8feed30a3d | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/10ff54a4-50e6-4550-b156-4209986768fc | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/613c98c1-2cc2-4bd1-b298-f3a9716c3755 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/55a3a888-558f-4866-b0ba-a5b430923d5a | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/e065d236-8360-4364-9d4e-91567e3706cd | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/e0d61fb9-8369-4fba-9e9c-09eeb9f682aa | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/24cffed0-8b54-45c6-8280-0f24fa6cbf21 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/b08e8dcf-397b-48ff-91eb-43dd9ad4523c | 880 | 59% | Bom |

Nesta categoria a maior pontuação atingida por um conjunto de dados foi 1000 pontos, atingindo 67% do máximo de pontuação, sendo classificado com a classificação de Bom. A menor pontuação atingida foi de 560 pontos, atingido 37% do máximo de pontuação, sendo classificado como Suficiente.

Na Figura 21, é possível observar o número e percentagem de conjuntos de dados encontrados para cada classificação.

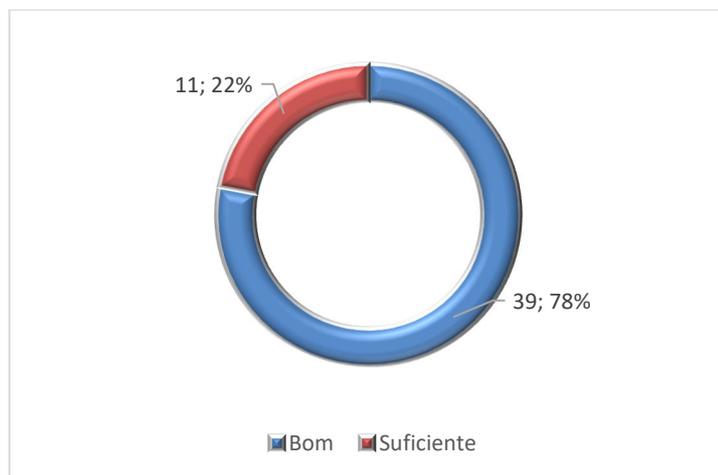


Figura 21 - Número e percentagem de conjuntos de dados por classificação, na categoria de Governo e Setor Público.

Na categoria de Governo e Setor Público, 39 dos conjuntos de dados têm a classificação de Bom e 11 conseguiram a classificação de Suficiente (Anexo 5).

Para a categoria de Transportes, os conjuntos de dados analisados segundo as DWBP estão descritos na Tabela 4.

Tabela 7 - Classificação dos conjuntos de dados na categoria de Transportes

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-aad3a2e382054549850fad6a6e907f29_0 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-c6c7530f6586438498be3b1d19955810_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-79fe4e405b044dcc91db3cf533d83a51_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-91ac370402d0426291ce71832bf9f9ef_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/021ce124-2ac0-4aee-a3d8-a1d33932b2ac | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-482a7fb892504d36bb2f7597fdcc9f6c_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/mangelmelder | 760 | 51% | Bom |
| https://data.europa.eu/data/datasets/rhein-kreis-neuss-2018-fahrzeugbestand | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/2194d38a-46a9-4e5c-8184-3e87ed1bfff8 | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/09870e46-26a3-4dc2-b632-4d1fba5092f9 | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/stanord cms-50476 | 940 | 63% | Bom |

Tabela 8 - Classificação dos conjuntos de dados na categoria de Transportes (Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-46131-012-regionalschluesel- | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-46241-221-regionalschluesel- | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-68c13623885b440ca1fa59249cb8d0a6_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-46251-001z-regionalschluesel- | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/hwtbh2wln9hggc bwh2ldg | 700 | 47% | Suficiente |
| https://data.europa.eu/data/datasets/stanord cms-62961 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/rhein-kreis-neuss-ladesaulen-in-deutschland | 740 | 49% | Suficiente |
| http://data.europa.eu/88u/dataset/JfuEDs5gCUQLRGCKnlZg | 700 | 47% | Suficiente |
| https://data.europa.eu/data/datasets/1127da4d-a1c3-44da-a8aa-526a24a46d50_1 | 680 | 45% | Suficiente |
| https://data.europa.eu/data/datasets/8f8a9f32-9247-43d9-8862-29e1c8946153 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/36e3596b-71d4-4bcb-bf03-040e662b5136 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/da57845f-ae51-41ca-944a-8f7beb132e8f?locale=pt | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/177e32de-cb85-4c1c-bb12-7188c6c49aa1 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/064a3764-84aa-48d0-ac43-f5b45f229584 | 1020 | 68% | Bom |
| https://data.europa.eu/data/datasets/8c953125-eec4-48c6-a562-416d36e7920f | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/573e82f1-83ca-406e-b126-31ee2558917a | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/fef6af4e-8ab0-48ef-96eb-182587f25f87 | 880 | 59% | Bom |
| https://data.europa.eu/data/datasets/5ca46dd0-a146-4666-b5f7-5fcfd869e7f7 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/a4cde4a6-6d8a-4627-817e-32449d48c34c | 940 | 63% | Bom |
| https://data.europa.eu/data/datasets/131e56d4-0a1d-4bd1-b9ab-306044d7647c | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/907d3a26-e79a-4009-a9cb-a294cbc44cab | 980 | 65% | Bom |
| https://data.europa.eu/data/datasets/c5b23223-59aa-4cee-8de3-da6e749e4a25 | 1000 | 67% | Bom |

Tabela 9 - Classificação dos conjuntos de dados na categoria de Transportes (Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/d8351c4a-bf41-4949-b108-5204f2973ee9 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/364fda21-fb6a-4415-a12c-9b876114bb85 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/4c99ef8f-ea58-444c-89a9-d480b274f1a7 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/5e425919-8818-4edb-96d2-839c6f4a3e8b | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/bd88bac4-6ba3-4cec-98d4-cd82083bdea3 | 760 | 51% | Bom |
| https://data.europa.eu/data/datasets/87345169-2ce4-4b24-831b-4dbb44cb6a5e | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/203218f4-00b3-4e8b-9deb-64b3b36661a0 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/6dbe7702-1644-4da8-bfb9-b612cfae7f1a | 980 | 65% | Bom |
| https://data.europa.eu/data/datasets/d147c582-6832-4419-abf4-08c5c4dcda9d | 940 | 63% | Bom |
| https://data.europa.eu/data/datasets/e4c29b67-f7ca-471f-9c62-0dfcf802d320 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/58918665-8cf2-4e8a-99c3-368d74f162af | 860 | 57% | Bom |
| http://data.europa.eu/88u/dataset/045h2PGTBBspSXEl79b0og | 600 | 40% | Suficiente |
| http://data.europa.eu/88u/dataset/g4PXFpgozszMwL&puTTg | 600 | 40% | Suficiente |
| https://data.europa.eu/data/datasets/3ba97e4d-fe68-409b-84e3-d312a7f45624_1 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/72099117-88f4-415e-9100-a471f2a878f3 | 680 | 45% | Suficiente |
| http://data.europa.eu/88u/dataset/6asaQpBYTpNftIm04psWTg | 600 | 40% | Suficiente |
| https://data.europa.eu/data/datasets/a2386e3c-c7f9-11e4-8731-1681e6b88ec1_1 | 780 | 52% | Bom |

Para a categoria de Transportes, a pontuação mais alta obtida por um conjunto de dados foi 1020 pontos atingindo 68% do máximo de pontuação, enquanto a mais baixa foi de 600 pontos, com um percentual de 40%.

Na Figura 22, é possível observar o número e percentagem de conjuntos de dados encontrados para cada classificação.

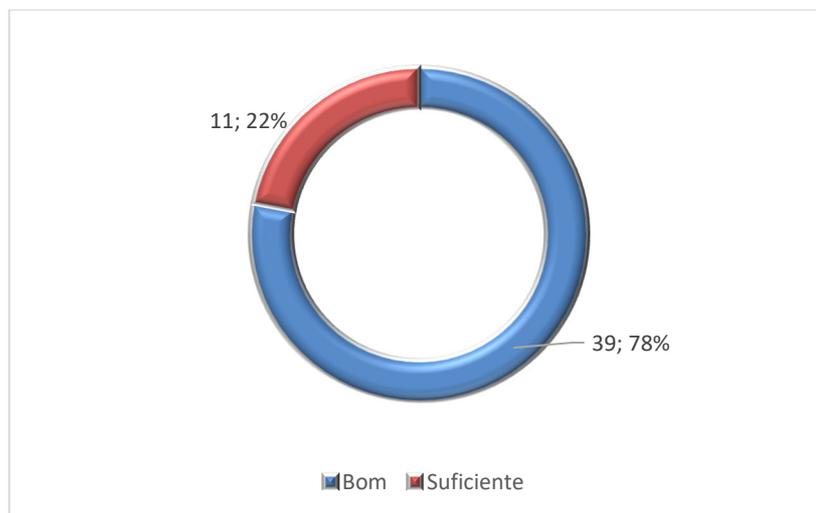


Figura 22 – Número e percentagem de conjuntos de dados por classificação, na categoria de Transportes.

Nesta categoria foram classificados como Bom, 39 conjuntos de dados, enquanto 11 foram classificados com Suficiente (Anexo 6).

Para a categoria de Educação, Cultura e Desporto, os conjuntos de dados analisados segundo as DWBP são apresentados na Tabela 5.

Tabela 10 - Classificação dos conjuntos de dados na categoria de Educação, Cultura e Desporto

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-2-5-33 | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-bild-55 | 640 | 43% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-0-22 | 720 | 48% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-2-4-4 | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-1-2-5 | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-bild-6 | 640 | 43% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-datenportal-bmbf-de-portal-2-5-27 | 680 | 45% | Suficiente |
| http://data.europa.eu/88u/dataset/yk3PpQjFzOQ51Zmjglg | 600 | 40% | Suficiente |
| https://data.europa.eu/data/datasets/2843fa46-ca6c-4098-b364-3f1491261d8d | 600 | 40% | Suficiente |
| http://data.europa.eu/88u/dataset/5lvoNLsAzD1njXsCGEkXA | 600 | 40% | Suficiente |
| https://data.europa.eu/data/datasets/stanord cms-49866 | 880 | 59% | Bom |

Tabela 11 - Classificação dos conjuntos de dados na categoria de Educação, Cultura e Desporto
(Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-21211-004z-regionalschluesel- | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-21111-206z-regionalschluesel- | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-8796b6f59dea4097865d9c1e3505c050_0 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-21111-204z-regionalschluesel- | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-3a5a260a54ec434e9abb3a3b7effd7a2_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/https-katalog-opendata-sachsen-de-store-3-resource-13 | 700 | 47% | Suficiente |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-1a7c930f08ce4628bb56bbc03ad07ff4_0 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-f9fdbbc098a42434ca9708872cc6b1854_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/stanord cms-62183 | 860 | 57% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-9f56609a638f45b79bdce651d8766835_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/227b95a4-f1f0-4188-b724-acb3eba36069 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-743b8b5c14e14d8ca9d8ead3894e3613_0 | 820 | 55% | Bom |
| http://portal-chemnitz.opendata.arcgis.com/datasets/a1a683ea78df4ac3b0d285588ba602f8_0 | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-07ba9bfbb0364a2e81d2457a1ea1c4aa_0 | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-statistik-sachsen-de-genonline-online-sequenz-tabelleergebnis-selectionname-21121-202z-regionalschluesel- | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/1b59b3e2-a01a-4b42-9a67-f133f66438c6 | 900 | 60% | Bom |

Tabela 12 - Classificação dos conjuntos de dados na categoria de Educação, Cultura e Desporto
(Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/rhein-kreis-neuss-bildung-und-teilhabe-but-anbieterverzeichnis | 760 | 51% | Bom |
| https://data.europa.eu/data/datasets/http-opendata-dortmund-de-duva2ckan-1-0-6-files-de-nrw-dortmund-studierende_der_tu_dortmund_insgesamt_und_nach_fakultaeten_im_wintersemester_seit_1990-11 | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/https-datos-alcobendas-org-dataset-agenda | 720 | 48% | Suficiente |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen65-c-40641-de | 740 | 49% | Suficiente |
| https://data.europa.eu/data/datasets/0f75d7ec-40ad-409e-8ad1-1a920559198f | 820 | 55% | Bom |
| https://data.europa.eu/data/datasets/https-datos-alcobendas-org-dataset-equipamiento-culturales-en-alcobendas | 720 | 48% | Suficiente |
| https://data.europa.eu/data/datasets/3d2fefc5-384a-4c1c-b55b-b55cfb34d321 | 900 | 60% | Bom |
| https://data.europa.eu/data/datasets/4fb2a687-4511-4523-8013-7b5235c71991 | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/f90b3f13-b3bf-4420-88d0-c6397efc62b9 | 780 | 52% | Bom |
| https://data.europa.eu/data/datasets/4b7c755f-d344-48c4-9aaf-0c72085428e5 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/https-opendata-euskadi-eus-catalogo-indicadores-municipales-de-sostenibilidad-poblacion-de-mas-de-10-anos-que-ha-completado-al-menos-estudios-secundarios- | 720 | 48% | Suficiente |
| https://data.europa.eu/data/datasets/bbd3b4a7-d375-4c72-9443-23de7026deff | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/2f123e1f-8e3c-4d2d-94c7-346317aef84 | 960 | 64% | Bom |
| https://data.europa.eu/data/datasets/eee67c82-3cc3-44f2-8f56-f0ad5c269233 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/ca6f89df-da06-4d5e-8a89-e0cd9d67ead1 | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/https-www-transparenz-bremen-de-de-datensatz-bremen65-c-40641-de | 840 | 56% | Bom |
| https://data.europa.eu/data/datasets/e44d7250-0b42-4d56-8713-7572a95aebb8 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/https-opendata-euskadi-eus-catalogo-directorio-de-recursos-culturales- | 720 | 48% | Suficiente |

Tabela 13 - Classificação dos conjuntos de dados na categoria de Educação, Cultura e Desporto
(Continuação)

| Identificador do conjunto de dados | Pontuação final | Percentual de pontos atingidos | Classificação |
|---|-----------------|--------------------------------|---------------|
| https://data.europa.eu/data/datasets/1509d3bf-acc1-4972-9447-bf0a201c51a6 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/f3d21de3-d78f-496d-81ed-b0046a383e10 | 760 | 51% | Bom |
| https://data.europa.eu/data/datasets/abc8961e-f5a4-4e21-870e-b37430a97201 | 1000 | 67% | Bom |
| https://data.europa.eu/data/datasets/d6f38430-255a-4456-b2ac-6c80652a957d | 800 | 53% | Bom |
| https://data.europa.eu/data/datasets/67784231-8ae4-4c89-bdc2-d6b76f4e3b7e | 880 | 59% | Bom |

Na categoria de Educação, Cultura e Desporto a pontuação mais alta atingida por um conjunto de dados foi de 1000 pontos, com um percentual de 67%. A pontuação mais baixa para um conjunto de dados nesta categoria foi de 600 pontos com um percentual de 40%.

Na Figura 23, é possível observar o número e percentagem de conjuntos de dados encontrados para cada classificação.

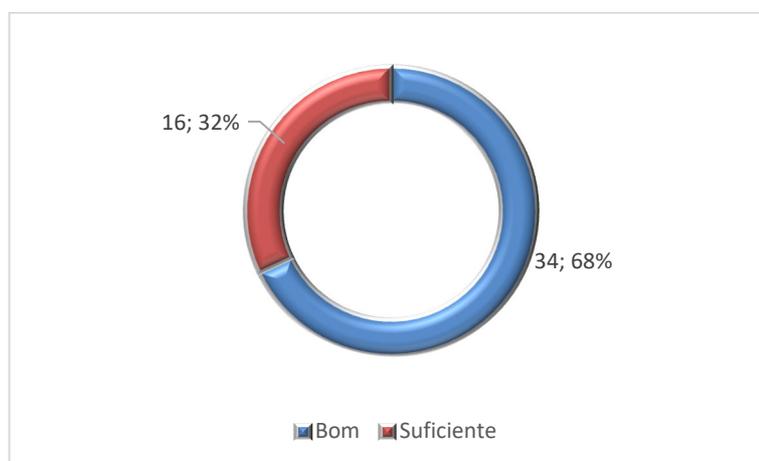


Figura 23- Número e percentagem de conjuntos de dados por classificação, na categoria de Educação, Cultura e Desporto.

Nesta categoria, foram encontrados 34 conjuntos de dados com a classificação de Bom e 16 com a classificação de Suficiente (Anexo 7).

No final da análise dos 150 conjuntos de dados relativos às três categorias, observou-se que 112 conjuntos de dados foram classificados, com base nas DWBP, como bons e 27 como suficientes. Quando confrontada com a avaliação feita pelo EDP, onde todos os conjuntos de

dados analisados tinham recebido a classificação de Bom+, percebe-se que a análise realizada segundo as DWBP, classificou alguns com uma classificação mais baixa (suficiente).

No entanto, vale salientar que este estudo não tem como objetivo comparar os resultados das avaliações realizadas pelo EDP com os resultados obtidos nesta investigação. A principal razão para que isso não seja viável é o facto de terem sido usados dois parâmetros distintos, os princípios FAIR, pelo EDP, e as Boas Práticas para Dados na Web – DWBP (W3C), por este estudo.

Esta questão fica clara no Quadro 1, que traz a síntese das correspondências dos indicadores adotadas pelo EDP, e tem por base os princípios FAIR, com as DWBP, onde nota-se que as DWBP são mais abrangentes. Essa abrangência reflete-se no facto de alguns aspetos não serem contemplados pelo Portal, como: verificar se os conjuntos de dados apresentam metadados de descrição; metadados estruturais; verificar se há controlo de versão de dados; verificar se os URI's são persistentes; contrariando as recomendações de Lócio et al. (2017) e Torino et al. (2019).

Como já referido, o EDP não avalia se os dados são disponibilizados através de uma API; se são disponibilizados mecanismos de *feedback* sobre os dados, o enriquecimento de dados; e se os conjuntos de dados citam a publicação original, elementos que melhoram a compreensão, a processabilidade, a capacidade de descoberta, a reutilização, a confiança, a capacidade de inter-ligação, o acesso e a interoperabilidade dos dados.

Assim, evidencia-se com este estudo que as DWBP, “concebidas para apoiar um ecossistema autossustentável”, que tem por objetivo facilitar a interação entre publicadores e consumidores (Lócio et al., 2017), ainda são pouco exploradas. Foi observada uma escassez de trabalhos sobre a adoção das DWBP, não só como orientações para publicação de dados, mas também como parâmetro para avaliação da qualidade de dados. Ressalta-se que a disponibilização dos dados não é suficiente, há que se assegurar a confiabilidade no acesso e no reuso.

5. CONCLUSÕES

Como referido inicialmente, o conceito de “dados” pode variar, consideravelmente, entre pesquisadores e diferentes áreas de conhecimento, uma vez que são gerados para diferentes fins, por diferentes comunidades e diferentes processos. Quando se refere a dados, é possível entender como sendo números, imagens, textos, vídeos ou *software*.

Está demonstrado pela literatura que os dados abertos auxilia as organizações a alcançar a Transformação Digital, a ter melhores custos económicos, e a melhorar a qualidade de vida dos cidadãos. No caso da utilização dos dados abertos no contexto dos transportes, auxilia a tomada de decisão por meio de diferentes tipos de visualização. Essas alternativas demonstram a importância da reutilização de dados.

Uma das iniciativas mundiais no uso dos dados abertos é o European Data Portal (EDP). O EDP tem como finalidade recolher metadados do setor público que se encontram disponíveis em portais de dados abertos de vários países Europeus. Para além disto, é também objetivo do EDP, promover a acessibilidade e valorizar os dados abertos.

Ao analisar o Portal, segundo as recomendações do W3C, foi possível responder a vários desafios encontrados na publicação ou consumo de dados na Web. As boas práticas recomendadas pela W3C para publicação de dados na Web respondem a desafios como o fornecimento de metadados, a licença dos dados, proveniência e qualidade, versão de dados, identificação de dados, formato de dados, vocabulário de dados, acesso aos dados, preservação de dados, *feedback*, enriquecimento de dados e republicação de dados.

Ao analisar as dimensões classificadas pelo EDP, foi possível concluir que algumas das DWBP são avaliadas por indicadores do EDP, sendo elas as seguintes: a BP1, BP2, BP4, BP9, BP10, BP11, BP12, BP14, BP15, BP21 e BP29.

Ao confrontar os resultados alcançados na avaliação dos conjuntos de dados segundo as DWBP, com os resultados atingidos pelo EDP, que adotam os princípios FAIR, chegou-se à conclusão de que, das DWBP mapeadas apenas 3 (BP1, BP4 e BP11) estão de acordo com os resultados do Portal.

De forma a classificar os conjuntos de dados avaliados à luz das DWBP, foi atribuído um peso igual a cada um dos benefícios alcançados pela utilização das BP's. Logo, a pontuação de cada BP deriva do número de benefícios alcançados pela mesma. A pontuação final de um conjunto de dados é a soma de cada uma das BP's presentes num conjunto de dados. A partir

dessa pontuação, consegue-se verificar em que medida os conjuntos de dados nas categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto, presentes no European Data Portal, cumprem as Boas Práticas para dados na Web identificadas pelo W3C, uma vez que, quanto maior a pontuação, maior é o número de boas práticas observadas pelo conjunto de dados.

Na análise dos conjuntos de dados, as BP's mais observadas foram as seguintes: BP1, BP2, BP4, BP5, BP8H, BP10, BP11, BP12, BP13C, BP16, BP17, BP19, BP29, BP32 e BP34. Enquanto, as menos observadas foram: BP3, BP6, BP7H, BP7C, BP8C, BP9, BP13H, BP14, BP15, BP21, BP22H, BP22C, BP30 e BP35.

Dependendo da pontuação final, os conjuntos de dados podem ser inseridos em 4 classificações: Excelente, Bom, Suficiente e Mau.

Na análise dos conjuntos de dados, na categoria de Governo e Setor Público foram encontrados 39 conjuntos de dados com a classificação de Bom, e 11 com a classificação de Suficiente. Na categoria de Transportes foram encontrados 39 conjuntos com a classificação de Bom e 11 com a classificação de Suficiente. Na categoria de Educação, Cultura e Desporto foram encontrados 34 com a classificação de Bom e 16 com a classificação de Suficiente.

Esses resultados revelam que, em certas situações, a avaliação dos conjuntos de dados realizada pelo Portal corresponde à desenvolvida à luz das Boas Práticas para Dados na Web. Em outras situações, as avaliações são diferentes, por vezes quase opostas às do EDP. Num trabalho futuro, pretende-se estender a análise para algumas BPs não contempladas neste estudo, a exemplo das relacionadas às APIs e à preservação; às demais categorias presentes no Portal; e elaborar recomendações específicas para os disponibilizadores de dados.

GLOSSÁRIO

De forma a contextualizar o tema em estudo desta dissertação, nesta subsecção são definidos os conceitos mais importantes e mais usados neste documento.

Dados

Pode-se entender como dados toda a informação ou gravações armazenadas de forma eletrónica, como por exemplo, documentos, bases de dados, transcrições e gravações áudio/visual (Makers, 2012). Quando os dados são públicos, estes não estão sujeitos a “privacidade válida, segurança ou limitações de privilégios” (Makers, 2012).

Metadados

Entende-se como metadados, dados sobre dados que estão presentes na Web. Os metadados podem incluir, mas não se limitando a: autoria, classificação, descrições, política, termos de distribuição e direitos autorais. Os metadados são bastante usados na Web Semântica (Berners-Lee, 1999).

Dados Abertos

Pode-se considerar como dados abertos, todos os dados que permitam a sua livre utilização, reutilização e redistribuição, sem restrições, legíveis por máquina, que estão disponíveis gratuitamente para o público sem nenhuma limitação (Attard et al., 2015).

Dados Abertos Governamentais

Os Dados Abertos Governamentais são dados que pertencem a um subconjunto dos Dados Abertos, referindo-se a dados relacionados com o governo que se encontram abertos ao público. Exemplos de dados governamentais podem ser: orçamentos, despesas, população, recenseamento geográfico, dados relacionados com o clima, poluição, etc (Attard et al., 2015).

Portal de Dados

Com o objetivo de abrir a informação do setor público (o movimento de dados abertos), é possível recolher e publicar os conjuntos de dados num portal de dados permitindo que estes funcionem como “balcão único” para os consumidores de dados, onde estes podem pesquisar dados reais e aceder a estes, de maneira a que possam interagir com os mesmos (Attard et al., 2015). A principal função dos portais de dados é a gestão de metadados dos conjuntos de dados (Attard et al., 2015).

Interface de Programação de Aplicações (API)

Protocolo que tem como função ser utilizado como interface por componentes de *software*, para que possam comunicar uns com os outros (Ubaldi, 2013).

Conjunto de Dados

Coleção de dados, normalmente apresentada sob a forma de tabela (Ubaldi, 2013).

Pesquisabilidade

O princípio da Pesquisabilidade afirma que os dados devem ser identificados, descritos e registados/indexados de forma clara. Para isso, aos conjuntos de dados deve ser atribuído um identificador único e persistente, as características dos dados devem ser sistematicamente especificadas, normalmente através de formatos padrão e estes devem ser armazenados/indexados num recurso público (Boeckhout et al., 2018).

Acessibilidade

Como acessibilidade, percebe-se que os conjuntos de dados devem ser acessíveis através de um procedimento de acesso, de preferência automatizado e bem definido. Para isso, deve ser criado um procedimento que permite a autenticação e autorização de acesso aos dados, bem como um procedimento para a recuperação dos dados, se necessário. Os metadados devem ser sempre acessíveis, mesmo quando os dados subjacentes não estejam mais disponíveis (Boeckhout et al., 2018).

Interoperabilidade

Segundo a definição da Comunidade Europeia, a interoperabilidade é vista como a capacidade que as organizações têm de interagir para objetivos que são mutuamente benéficos, partilhando informação e conhecimento entre organizações através de processos de negócio e troca de dados entre os seus sistemas TIC (Union, 2004).

Como o princípio da interoperabilidade, entende-se que os “dados e metadados sejam concetualizados, expressos e estruturados utilizando normas comuns, publicadas” (Boeckhout et al., 2018). Para que este princípio de verifique, é necessário utilizar formatos de dados padrão, variáveis, ontologias, entre outras coisas.

Reusabilidade

O princípio da reusabilidade estipula que os dados devem apresentar as suas características, bem como a sua proveniência, de forma clara e ao pormenor de acordo com as normas relevantes, de forma clara e acessível para utilização. Para isto, os conjuntos de dados devem apresentar descrições exatas dos dados, licenças de acesso e utilização, normas que estiveram subjacentes ao processo e toda a proveniência associada ao conjunto de dados (Boeckhout et al., 2018).

Contextualidade

A dimensão contextualidade permite ao utilizador obter mais informações sobre o conjunto de dados, de maneira que este fique mais contextualizado sobre o mesmo. Para esta dimensão é possível obter informações, por exemplo, sobre os direitos dos conjuntos de dados, o tamanho do ficheiro, a data de emissão e a data de emissão (Data.europa.eu, 2021).

BIBLIOGRAFIA

- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Berends, J., Carrara, W., & Radu, C. (2017). *Analytical Report 9: The Economic Benefits of Open Data*. <https://doi.org/10.2830/081754>
- Berends, J., Carrara, W., Vollers, H., Fechner, T., & Kleemann, M. (2017). *Analytical Report 5: Barriers in working with Open Data*, 37. Retrieved from http://www.europeandataportal.eu/%0Ahttps://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf
- Berends, J., Carrara, W., & Vollers, H. (2020). *Analytical Report 6: Open Data in Cities 2*. <https://doi.org/10.2830/007197>
- Berners-Lee, T. Weaving the Web. *Glossary* (1999). Retrieved May 5, 2021, from <https://www.w3.org/People/Berners-Lee/Weaving/glossary.html>
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018, July 1). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*. Nature Publishing Group. <https://doi.org/10.1038/s41431-018-0160-0>
- Burle, C. (2021). *As Boas Práticas para Dados na Web*. Retrieved from <https://acervo.ceweb.br/acervos/conteudo/43a2f21d-56be-4798-a10a-e0a7a054a53c>.
- Carrara, W., Engbers, W., Nieuwenhuis, M., & Steenbergen, E. van. (2020). *Analytical report 4 Open data in cities* (pp. 1–28). Retrieved from https://data.europa.eu/sites/default/files/edp_analytical_report_n4_-_open_data_in_cities_v1.0_final.pdf
- Carrara, W., Fischer, S., Oudkerk, F., Steenbergen, E. van, & Tinholt, D. (2015). *Analytical Report 1: Digital Transformation and Open Data*. 1–22. <https://doi.org/10.2830/673557>
- Carrara, W., Fischer, S., & Steenbergen, E. van. (2020). *Analytical Report 2: E-skills and Open Data*. Retrieved from https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n2_-_e-skills.pdf
- Dados.gov.br. (n.d.). *Produto GT 2 - Modelo de Maturidade de publicação de dados - Infraestrutura Nacional de Dados Abertos (INDA)*. Retrieved May 31, 2021, from <https://wiki.dados.gov.br/Produto-GT-2-Modelo-de-Maturidade.ashx>
- Data.Europa.Eu. (2021 update). About Data.Europa.Eu. Retrieved from <https://data.europa.eu/de/highlights/open-regions-and-cities-data-european-data-portal>
- Data.Europa.EU. (2021). Portal oficial dos dados europeus. *Conjuntos de dados*. Retrieved January 31, 2021, from <https://www.europeandataportal.eu/data/datasets?locale=pt&minScoring=0&page=1.>
- Data.europa.EU. (2021). *Metodologia: painel de avaliação de metadados*. (n.d.). Retrieved January 21, 2021, from <https://www.europeandataportal.eu/mqa/methodology?locale=pt>
- Dunning, A., De Smaele, M., & Böhmer, J. (1970). Are the FAIR Data Principles fair? *International Journal of Digital Curation*, 12(2), 177–195. <https://doi.org/10.2218/ijdc.v12i2.567>
- European Commission. (2021). *Open data portals*. Retrieved May 28, 2021, from <https://digital-strategy.ec.europa.eu/en/policies/open-data-portals>

- European Data Portal. (2020). *Open Data Maturity*. <https://doi.org/10.2830/073835>
- European Union. (2004). *New European Interoperability Framework: Promoting seamless services and data flows for European public administrations*. White Pages (pp. 1–40). Publications Office of the European Union. Retrieved from https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf
- For, G., Of, E., Jos, M., Sim, M. G., Farias, S. H., Sim, R. G., & Federal, U. (2009). *Metodologia da pesquisa científica: diretrizes para a elaboração de um protocolo de pesquisa 1*.
- FORCE11. The Future of Research Communications and e-Scholarship (2015). *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0 | FORCE11*. <https://www.force11.org/fairprinciples>
- Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data : Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4), 441–452. <https://doi.org/10.1016/j.giq.2015.09.001>
- Huijboom, N., & Broek, T. Van Den. (2011 April). Open data : an international comparison of strategies. *European Journal of ePractice*, 1–13.
- Huyer, E. (2018). *Analytical Report 11: Re-use of PSI in the public sector Analytical Report n 11*. Retrieved from <http://www.europeandataportal.eu/>
- Huyer, E., & Blank, M. (2015). *Analytical Report 15: High-value datasets: understanding the perspective of data providers*. 1–6. <https://doi.org/10.2830/363773>
- Huyer, E., & Cecconi, G. (2020). *Analytical Report 12: Business-to-Government Data Sharing*. 1–24. <https://doi.org/10.2830/078126>
- Kawashita, I., Baptista, A. A., & Soares, D. S. (2020). An Assessment of Open Benchmark Instruments Government. In: International Conference EGOV-CeDEM-ePart 2020, Linköping, 2020. *CEUR Workshop Proceedings*. Aachen: CEUR-WS. p. 29-40.
- Knippenberg, L. van. (2015). *Analytical Report 16: Open Data Best Practices in Europe: Learning From Cyprus, France, and Ireland*. <https://doi.org/10.2830/937861>
- Lakatos, E.M., Marconi, M.A. (2003). *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas).
- Lóscio, B.F., Burle, C., Calegari, N. (eds.). *Data on the Web Best Practices*. (2017, January 31). <https://www.w3.org/TR/dwbp/#intro>
- Lóscio, B. F., Burle, C., & Calegari, N. (2019). Implementation of the best practices for data on the web in Brazil and Costa Rica. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, 807–810. <https://doi.org/10.1145/3308560.3316463>
- Makers, D. (2012). *Linked Open Data : The Essentials*. Retrieved from <http://www.semantic-web.at/LOD-TheEssentials.pdf>
- Murray-Rust, P. (2008). Open data in science. *Serials Review*, 34(1), 52–64. <https://doi.org/10.1080/00987913.2008.10765152>
- National Spatial Data Infrastructure. About nipp.hr. 2021. Retrieved from <https://www.nipp.hr/default.aspx?id=1728>
- Oliveira, L. E. R. A., Oliveira, M. I. S., & Lóscio, B. F. (2017). *Um Survey sobre Soluções para Publicação de Dados na Web sob a Perspectiva das Boas Práticas do W3C*. 148–159.
- Oliveira, L. E. R. A., Oliveira, M. I. S., Do Rego Santos, W. C., & Lóscio, B. F. (2018). Data on the web management system: A reference model. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3209281.3209355>
- Paschoarelli, L. C., Medola, F. O., Henrique, G., & Bonfim, C. (2015). *Características Qualitativas, Quantitativas e Quali- quantitativas de Abordagens Científicas : estudos de*

- caso na subárea do Design Ergonômico*. *Revista Do Design, Tecnologia e Sociedade*, 2(1), 65–78.
- Penteado, B. E., Bittencourt, I. I., & Isotani, S. (2019). Análise exploratória sobre a abertura de dados educacionais no Brasil: como torná-los prontos para o ecossistema da Web? *Revista Brasileira de Informática Na Educação*, 27(01), 175. <https://doi.org/10.5753/rbie.2019.27.01.175>
- Priego, L. P., & Osimo, D. (2020). *Analytical Report 17 From Regulation To Adoption: The Drivers Of Data Sharing*. <https://doi.org/10.2830/850048>
- Radu, C. (2020). *Analytical Report 13: Open Data best practices in Europe's Top Performers: Ireland, Spain and France*. <https://doi.org/10.2830/05271>
- Raupp, F. M., & Beuren, I. M. (2013). *Metodologia da pesquisa aplicável às ciências sociais*. São Paulo: Atlas.
- Sayão LF, Sales, LF. Dados abertos de pesquisa : ampliando o conceito de acesso livre Open research data : extending the concept of free access. *RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde*, 2014, 8(2), 76–92. Retrieved from <https://www.arca.fiocruz.br/handle/icict/17102>. doi: 10.3395/reciis.v8i2.934.pt.Simperl, E., O'Hara, K., & Gomer, R. (2020). *Analytical Report 3: Open Data and Privacy*. <https://doi.org/10.2830/532195>
- Simperl, E., & Walker, J. (2017). *Analytical Report 8: The Future of Open Data Portals*, 1–26. Retrieved from https://www.europeandataportal.eu/sites/default/files/edp_analyticalreport_n8.pdf
- Torino, E., Trevisan, G. L., Vidotti, S.A.B.G. (2019). Dados abertos CAPES: um olhar à luz dos desafios para publicação de dados na Web. *Ci. Inf.* 48(Supl.):38-46.
- Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, NO.22(22), 61. <http://www.oecd-ilibrary.org/docserver/download/5k46bj4f03s7.pdf?expires=1400038714&id=id&accname=guest&checksum=0A581F24362C40FF67A6EF59C2B31BBB>
- Walker, J., & Simperl, E. (2020). *Analytical Report 10: Open Data and Entrepreneurship*. 1–22. <https://doi.org/10.2830/357754>
- Zeng, M. L., Hlava, M., Greenberg, J., & Needleman, M. (2017). Standards and best practices related to the publication, exchange and usage of open data. *Proceedings of the Association for Information Science and Technology*, 54(1), 610–611. <https://doi.org/10.1002/pr2.2017.14505401086>

ÍNDICE DE APÊNDICES

Anexo 1 – Estudo exploratório

Anexo 2 – Estudo final – Governo e setor público

Anexo 3 – Estudo final – Transportes

Anexo 4 – Estudo final – Educação, Cultura e Desporto

Anexo 5 – Classificação Governo e Setor Público

Anexo 6 – Classificação Transportes

Anexo 7 – Classificação Educação, Cultura e Desporto

Nas tabelas, as Boas práticas foram identificadas por BP e as descrições das abreviações referem-se a:

BP1- Fornecer metadados

BP2- Fornecer metadados descritivos

BP3- Fornecer metadados estruturais

BP4- Fornecer informações sobre licença de dados

BP5- Fornecer informações de proveniência de dados

BP6- Fornecer informações de qualidade de dados

BP7- Fornecer um indicador de versão

BP8- Fornecer histórico de versões

BP9- Usar URIs persistentes como identificadores de conjuntos de dados

BP10- Usar URIs persistentes como identificadores dentro de conjuntos de dados

BP11- Atribuir URIs a versões e séries de conjuntos de dados

BP12- Usar formatos de dados padronizados legíveis por máquina

BP13- Usar representações de dados neutros

BP14- Fornecer dados em vários formatos

BP15- Reutilizar vocabulários, preferencialmente padronizados

BP16- Escolher o nível certo de formalização

BP17- Fornecer download em massa

BP18- Fornecer subconjuntos para grandes conjuntos de dados

BP19- Usar negociação de conteúdo para servir dados disponíveis em vários formatos

- BP20- Fornecer acesso em tempo real
- BP21- Fornecer dados atualizados
- BP22- Fornecer uma explicação para dados que não estão disponíveis
- BP23- Disponibilizar dados através de uma API
- BP24- Usar padrões da web como base de APIs
- BP25- Fornecer documentação completa para a API
- BP26- Evitar quebrar alterações na API
- BP27- Preservar identificadores
- BP28- Avaliar a cobertura do conjunto de dados
- BP29- Reunir *feedbacks* dos consumidores de dados
- BP30- Disponibilizar *feedback*
- BP31- Enriquecer dados gerando novos dados complementares
- BP32- Proporcionar apresentações complementares
- BP33- Fornecer *feedback* ao editor original
- BP34- Seguir os Termos de Licenciamento
- BP35- Citar a publicação original

ANEXO 1- ESTUDO EXPLORATÓRIO

O estudo exploratório incidu sobre as categorias de Governo e Setor Público, Transportes e Educação, Cultura e Desporto e permitiu validar a técnica de amostragem definida inicialmente. A técnica de amostragem inicial passou por analisar os 20 primeiros conjuntos de dados de cada categoria, ordenados pela última modificação e filtrados pela classificação de excelente ou bom+.

Nas linhas do estudo exploratório encontram-se as identificações dos conjuntos de dados das três categorias e nas colunas as melhores práticas. Quando o conjunto de dados em análise atende a melhor prática foi colocado a abreviação S, caso não tenha sido atendido colocou-se N e quando não foi identificado colocou-se NI. Em algumas BP's foi colocado a abreviação NA, o que significa que não se aplica.

O estudo exploratório revelou que a técnica de amostragem utilizada necessitava de ser melhorada, uma vez que, os conjuntos analisados neste estudo referiam-se muitas vezes ao mesmo tema, alterando apenas o ano a que se referiam os dados, o que tornava a análise muito limitada.

ANEXO 1- ESTUDO EXPLORATÓRIO

Tabela 1 – Análise preliminar dos conjuntos de dados das categorias Governo e Setor Público, Transportes e Educação, Cultura e Desporto

| Práticas | BP1 | BP2 | BP3 | BP4 | BP5 | BP6 | BP7 | BP8 | BP9 | BP10 | BP11 | BP12 | BP13 | BP14 | BP15 | BP16 | BP17 | BP18 | BP19 | BP20 | BP21 | BP22 | BP23 | BP24 | BP25 | BP26 | BP27 | BP28 | BP29 | BP30 | BP31 | BP32 | BP33 | BP34 | BP35 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Governo e setor público (excelente) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 008b0185-a5b3-4b70-af94-617f692760fa | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| fe933fc6-3c0d-44a5-bd14-75d79a50b0c0 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| c276228b-9c7b-4812-b4a5-7480316c2233 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| 9428cf36-a3a9-40bc-9d82-3964ae497c9b | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| bdbbae58-41e0-4453-bfe3-369e5c427cd0 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| f508e9f3-17db-4bb1-916b-bdbb9295bde0 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| 655f46c2-102b-45a6-a9c4-cc756d47df | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| d1b9d1eb-4d62-4996-ae7e-281195d3b766 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| 28614843-8f98-4ef5-8ac4-4956ae2455ba | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| d02483e4-6dca-45ee-9cf4-97cf27742179 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| c965a646-70fd-46e9-8ffb-b3254f9a3a4b | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| 7048ddc7-c9c1-4017-bef6-b679e243631b | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| 15c1277f-5b7b-445e-a7c0-a80c714c042a | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| c2f73bc6-02dd-48ae-b9c6-4b8fe14da580 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| a3ccc8c3-810d-49c2-a1c2-b82645e87c97 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| d418059b-0e60-4995-91e9-7a9a17f4b3aa | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| b2e2f9be-b977-4459-a9b0-e1bf303b0d99 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| e2ae83b6-9b9b-4f07-8399-35442395299c | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| d31b7c96-c44a-46d3-87b7-e6275d7637ca | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| ef69aa52-2315-4ec4-a62c-a625834e5fa9 | S | S | NI | S | S | S | S | S | N | S | S | S | N | S | Parc.S | S | S | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| Transportes (BOM+) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| https://www.europeandataportal.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-aad3a2e382054549850fad6a6e907f9_0?locale=pt | S | S | NI | S | S | S | N | N | N | N | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/482a7fb892504d36bb2f759... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | NI | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/68c13623885b440ca1fa592... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | NI | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/c6c7530f6586438498be3b1... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | NI | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/79fe4e405b044cc91db3c... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | NI | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/91ac370402d0426291ce718... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | N | NI | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/stanord_cms-62581?locale=pt | S | S | NI | S | S | S | S | N | N | NI | S | S | N | N | NI | N | NA | N | NA | S | N | NA | S | N | NA | N | NA | S | S |
| https://www.europeandataportal.eu/data/datasets/c64840bb-80bd-43a4-974a-7ea78025b059?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | N | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/4ff7f6b6-9ffe-4b89-93ef-0be591ef931d?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | N | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-katalog-opendata-sachsen-de-store-6-resource-11?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | N | NA | S | N | NA | S | N | NA | N | NA | S | N |
| mangelmelder | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| rhein-kreis-neuss-2018-fahrzeugbestand | S | S | NI | S | S | S | S | N | N | NI | S | S | N | S | N | NI | N | NA | S | NA | S | N | NA | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | S |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-scharbeutz_strandabschnitt_9-13_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-scharbeutz_strandabschnitt_21-24_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-parkplatz_scharbeutz_p3_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-parkplatz_scharbeutz_p2_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-parkplatz_haffkrug_p1_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-parkplatz_scharbeutz_badeweg_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-scharbeutz_strandabschnitt_14-20_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/https-opendata-schleswig-holstein-de-dataset-parkplatz_scharbeutz_parkbuechten_strandallee_ii_2021-01-03?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | NI | N | NA | NI | NA | S | N | NA | N | N | NA | N | NA | S | N |
| Educação, Cultura e Desporto (Excelente) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7afdfa7a-7fd4-48a2-be63-e83750b1c40 | S | S | NI | S | S | S | S | N | N | S | S | S | N | S | Parc.S | NI | S | NA | S | NA | N | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | S |
| http://portal-chemnitz.opendata.arcgis.com/datasets/8796b6f59dea4097865d9c1... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| https://www.europeandataportal.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-07ba9bfb0364a2e81d2457a1ea1c4a_0?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/9f56609a638f45b79bdce65... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/datasets/a1a683ea78df4ac3b0d2855... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N |
| http://portal-chemnitz.opendata.arcgis.com/d | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| Práticas | BP1 | BP2 | BP3 | BP4 | BP5 | BP6 | BP7 | BP8 | BP9 | BP10 | BP11 | BP12 | BP13 | BP14 | BP15 | BP16 | BP17 | BP18 | BP19 | BP20 | BP21 | BP22 | BP23 | BP24 | BP25 | BP26 | BP27 | BP28 | BP29 | BP30 | BP31 | BP32 | BP33 | BP34 | BP35 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| https://www.europeandataportal.eu/data/datasets/https-katalog-opendata-sachsen-de-store-3-resource-13?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | N | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| http://portal-chemnitz.opendata.arcgis.com/datasets/5fc83eb0ff484bc59eb50e6... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| http://portal-chemnitz.opendata.arcgis.com/datasets/1f7ed871eabe412b828fa4b... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| http://portal-chemnitz.opendata.arcgis.com/datasets/f9fdb098a42434ca970887... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| http://portal-chemnitz.opendata.arcgis.com/datasets/743b8b5c14e14d8ca9d8ead... | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| https://www.europeandataportal.eu/data/datasets/http-portal-chemnitz-opendata-arcgis-com-datasets-6a7c8e18637d49d8aa29f69f4106f596_0?locale=pt | S | S | NI | S | S | S | N | N | N | NI | S | S | N | S | N | NI | S | NA | N | NA | S | N | NA | NA | NA | NA | NA | S | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/1.8.2 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/0.3 | S | S | NI | S | S | S | S | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/0.29 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/0.59 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/2.1.14 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/2.5.106 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |
| https://www.datenportal.bmbf.de/portal/0.56 | S | S | NI | S | S | S | N | N | S | NI | S | S | N | S | N | NI | N | NA | N | NA | S | N | NA | NA | NA | NA | NA | N | N | NA | N | NA | S | N | |

Nota: Parc. = Parc.

ANEXO 2- ESTUDO FINAL GOVERNO E SETOR PÚBLICO

No estudo final da categoria de Governo e Setor Público, foram analisados os 50 primeiros conjuntos de dados de acordo com a técnica de amostragem por conveniência. Esta técnica permite à análise dos conjuntos de dados verificar se o conjunto em tela pertence à mesma categoria do anterior, alterando por exemplo apenas o ano a que se referem os dados, e se isto acontecer passa-se à frente, até completar os 50 conjuntos de dados.

Nas linhas das tabelas encontram-se as BP's, bem como uma linha de observações para cada uma. Nas colunas encontram-se os identificadores dos conjuntos de dados analisados. Caso o conjunto de dados em análise observe a melhor prática foi colocado a abreviação S, caso não observe colocou-se N e quando não foi identificado colocou-se NI. Em algumas BP's foi colocado a abreviação NA, o que significa que não se aplica.

Quando é necessário fazer alguma observação sobre o conjunto de dados em relação a alguma BP, utilizou-se o campo observações.

ANEXO 2- ESTUDO FINAL GOVERNO E SETOR PÚBLICO

Tabela 1 – Análise final da categoria Governo e Setor Público (Bom+)

| BEST PRACTICES | LEVEL | URL | BP1-Remarks | BP2-Remarks | BP3-Remarks | BP4-Remarks | BP5-Remarks | BP6-Remarks | BP7-Remarks | BP8-Remarks | BP9-Remarks | BP10-Remarks | BP11-Remarks |
|----------------|-------|---|-------------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|--------------|
| | | https://data.europa.eu/data/datasets/eu-whoswho-the-official-directory-of-the-european-union?locale=pt | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/covid-19-multilingual-terminology | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/figs | Y | Does not show creation date and modification date | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/official-journals-of-the-european-union-2011_14 | Y | | | | | | | | | | |
| | | https://europeandataportal.eu/sets/data/https-datosabiertos-una-animada-es-datos-abiertos-de-accion-municipal-leon-leon | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/public-opinion-in-the-eu-in-time-of-coronavirus-crisis-2 | Y | Does not show creation date and modification date | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/eu-open-data-portal-all-datasets-and-their-metadata | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/5Xvfr7NouitYs514v7A | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/h1eq482p65647mvdnsg | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/QE0oTuzZfVwZy62lg | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/YsRgIGg1veYELUuWj9BA | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/Qavm4mN6ZyvasfQ5Q | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/OEJU0Bj20MeH2p5GFObw | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/XZeiRPrWRdGnjyHjEOXw | Y | Does not show creation date | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/class-sum-leg | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/file-status | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/legal-act-domain | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/role-qualifier | Y | | | | | | | | | | |
| | | http://data.europa.eu/88u/dataset/european-commission-hr-key-figures-2016 | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/https-www-transparentz-bremen-de-de-datensatz-bremen53--81168-de?locale=fr | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/https-www-transparentz-bremen-de-de-datensatz-bremen236-c-8202-de?locale=fr | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/https-www-transparentz-bremen-de-de-datensatz-bremen53-c-62365-de?locale=fr | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/https-www-transparentz-bremen-de-de-datensatz-bremen53--63348-de?locale=fr | Y | | | | | | | | | | |
| | | https://europeandataportal.eu/sets/data/https-www-transparentz-bremen-de-de-datensatz-bremen53-c-81168-de | Y | | | | | | | | | | |
| | | https://europeandataportal.eu/sets/data/https-www-transparentz-bremen-de-de-datensatz-bremen53-c-76143-de | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/https-www-transparentz-bremen-de-de-datensatz-bremen236-c-8249-de?locale=ga | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/a7dca82-bed4-4db2-9767-59dc42dbac3a | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/845089c0-2ec3-4ce4-b074-118825d8e1d1 | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/b1ef48f-601f-4a08-844a-444444444444 | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/c55386e0-5bc0-4b89-844a-444444444444 | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | https://data.europa.eu/data/datasets/http-portal-chemnitz | Y | | | | | | | | | | |
| | | | | | | | | | | | | | |

Nota: S- Sim// N- Não// NA- Não aplicado

ANEXO 3- ESTUDO FINAL TRANSPORTES

No estudo final da categoria de Transportes, foram analisados os 50 primeiros conjuntos de dados de acordo com a técnica de amostragem por conveniência. Esta técnica permite na análise aos conjuntos de dados verificar se o conjunto em análise pertence à mesma categoria do anterior, alterando por exemplo apenas o ano a que se referem os dados, e se isto acontecer passa-se à frente, até completar os 50 conjuntos de dados.

Nas linhas das tabelas encontram-se as BP's, bem como uma linha de observações para cada uma. Nas colunas encontram-se os identificadores dos conjuntos de dados analisados. Caso o conjunto de dados em análise observe a melhor prática, é colocado a abreviação S, caso não observe colocou-se N e quando não foi identificado colocou-se NI. Em algumas BP's foi colocado a abreviação NA, o que significa que não se aplica.

Quando é necessário fazer alguma observação sobre o conjunto de dados em relação a alguma BP, utilizou-se o campo observações criado para o efeito.

Anexo 4- Estudo Final Educação, Cultura E Desporto

No estudo final da categoria de Educação, Cultura e Desporto, foram analisados os 50 primeiros conjuntos de dados de acordo com a técnica de amostragem por conveniência. Esta técnica permite na análise aos conjuntos de dados verificar se o conjunto em análise pertence à mesma categoria do anterior, alterando por exemplo apenas o ano a que se referem os dados, e se isto acontecer passa-se à frente, até completar os 50 conjuntos de dados.

Nas linhas das tabelas encontram-se as BP's bem como uma linha de observações para cada uma. Nas colunas encontram-se os identificadores dos conjuntos de dados analisados. Caso o conjunto de dados em análise observe a melhor prática foi colocado a abreviação S, caso não observe colocou-se N e quando não foi identificado colocou-se NI. Em algumas BP's foi colocado a abreviação NA, o que significa que não se aplica.

Quando é necessário fazer alguma observação sobre o conjunto de dados em relação a alguma BP, utilizou-se o campo observações criado para o efeito.

ANEXO 5- CLASSIFICAÇÃO GOVERNO E SETOR PÚBLICO

A classificação da categoria de Governo e Setor Público permitiu aferir a pontuação final de cada conjunto de dados analisado para esta categoria. Para além da pontuação final, o conjunto de dados também foi classificado quanto ao percentual de pontos atingido em relação à pontuação máxima que um conjunto de dados conseguia obter, bem como a sua classificação qualitativa.

Para cada BP atendida foi associado o número um, em caso negativo, não identificado ou não aplicado, foi associado zero. Para cada BP presente foi efetuado a soma de cada uma, sendo que a soma total corresponde à pontuação final de cada conjunto de dados. Depois de obtida a pontuação final, esta foi convertida em percentagem e de acordo com a percentagem alcançada, foi realizada a classificação qualitativa de cada conjunto de dados.

ANEXO 6- CLASSIFICAÇÃO TRANSPORTES

A classificação da categoria de Transportes permitiu aferir a pontuação final de cada conjunto de dados analisado para esta categoria. Para além da pontuação final, o conjunto de dados também foi classificado quanto ao seu percentual de pontos atingido em relação à pontuação máxima que um conjunto de dados conseguia obter, bem como a sua classificação qualitativa.

Para cada BP atendida foi associado o número um, em caso negativo, não identificado ou não aplicado, foi associado zero. Para cada BP presente foi efetuado a soma de cada uma, sendo que a soma total corresponde à pontuação final de cada conjunto de dados. Depois de obtida a pontuação final, esta foi convertida em percentagem e de acordo com a percentagem alcançada foi realizada a classificação qualitativa de cada conjunto de dados.

ANEXO 7- CLASSIFICAÇÃO EDUCAÇÃO, CULTURA E DESPORTO

A classificação da categoria de Educação, Cultura e Desporto permitiu aferir a pontuação final de cada conjunto de dados analisado para esta categoria. Para além da pontuação final, o conjunto de dados também foi classificado quanto ao percentual de pontos atingido em relação à pontuação máxima que um conjunto de dados conseguia obter, bem como a sua classificação qualitativa.

Para cada BP atendida foi associado o número um, em caso negativo, não identificado ou não aplicado, foi associado zero. Para cada BP presente foi efetuado a soma unitária, sendo que a soma total corresponde à pontuação final de cada conjunto de dados. Depois de obtida a pontuação final, esta foi convertida em percentagem, e de acordo com a percentagem alcançada, foi realizada a classificação qualitativa de cada conjunto de dados.

