



**A UTILIZAÇÃO DE DADOS PÚBLICOS PARA A ANÁLISE DE DADOS:
UMA PESQUISA BIBLIOGRÁFICA COM A UTILIZAÇÃO DE DADOS
GOVERNAMENTAIS EM FORMATO ABERTO**

Cláudio Sonáglio Albano – claudioalbano@unipampa.edu.br
UNIVERSIDADE FEDERAL DO PAMPA – UNIPAMPA

Manrik Santos Souza – manriksouza.aluno@unipampa.edu.br
UNIVERSIDADE FEDERAL DO PAMPA – UNIPAMPA

ÁREA: 06. ENGENHARIA ORGANIZACIONAL
SUBÁREA: 06.4 - GESTÃO DA INFORMAÇÃO

RESUMO: EXISTE UMA TENDÊNCIA PELA DISPONIBILIZAÇÃO DE DADOS PÚBLICOS (DE GOVERNOS) EM FORMATO ABERTO, USANDO OS RECURSOS DA INTERNET. EM PARALELO É CRESCENTE O USO DE CONCEITOS E FERRAMENTAS DA ÁREA DE CIÊNCIA DE DADOS VISANDO MELHOR SUPORTE PARA O PROCESSO DE TOMADA DE DECISÃO. ASSIM, ESTE TRABALHO TEVE COMO OBJETIVO GERAL VERIFICAR A POSSIBILIDADE DA UTILIZAÇÃO DE DADOS PÚBLICOS, EM FORMATO ABERTO, FAZENDO USO DAS POTENCIALIDADES DA CIÊNCIA DE DADOS, DE FORMA A MELHORAR A GESTÃO PÚBLICA, VISANDO AVANÇAR NO ENTENDIMENTO, NA APLICAÇÃO PRÁTICA E NA PROMOÇÃO DE INOVAÇÕES COLABORATIVAS. POR INTERMÉDIO DE UMA REVISÃO BIBLIOGRÁFICA EM QUATRO BASE DE DADOS, FORAM PESQUISADOS TRABALHOS QUE ABORDARAM A UTILIZAÇÃO DE DADOS PÚBLICOS EM FORMATO ABERTO PARA ANÁLISE DE DADOS, UTILIZANDO CONCEITOS DA CIÊNCIA DE DADOS. OS PRINCIPAIS RESULTADOS DEMONSTRAM QUE: EXISTE UMA QUANTIDADE SIGNIFICATIVA DE ESTUDOS RELACIONADOS À UTILIZAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS PARA ANÁLISE DE DADOS, ABRANGENDO DIVERSAS ÁREAS COMO SAÚDE, EDUCAÇÃO E SEGURANÇA. ESTE FATO DEMONSTRA O POTENCIAL DESSES DADOS PARA O SUPORTE A DECISÕES, EM BUSCA DE MELHORES RESULTADOS E MAIOR PROMOÇÃO DE INOVAÇÃO ABERTA, PELA POSSIBILIDADE DOS DADOS SEREM USADOS POR TODA A SOCIEDADE.

PALAVRAS-CHAVES: CIÊNCIA DE DADOS; DADOS PÚBLICOS; DADOS ABERTOS; TOMADA DE DECISÃO.

.

;

1. INTRODUÇÃO

A ciência de dados desponta como um campo interdisciplinar que cada vez ganha mais reconhecimento por sua capacidade de facilitar a aplicação de métodos científicos na extração de *insights* para orientar decisões. Os cientistas de dados exploram uma variedade de domínios, incluindo matemática, estatística, ciência da computação, engenharia de software, aprendizado de máquina, análise de dados, mineração de dados e textos, gestão de projetos e *big data*, navegando por um amplo espectro de conhecimento para criar impacto (CALDEIRA, 2022).

Para Teixeira (2003) o processo de coleta e análise de dados se desenha de maneira flexível e em constante evolução, destacando-se por sua abordagem interativa e dinâmica. A análise de dados é o caminho para extrair significado para além dos dados brutos, é um processo que envolve consolidar, delimitar e interpretar o que as pessoas expressaram e o que o pesquisador observou e leu, resultando na formação de significados mais amplos.

Em virtude de diversos recursos tecnológicos, qualquer pessoa tem a possibilidade de coletar, armazenar, divulgar e utilizar dados e/ou informações sujeitas apenas a requisitos mínimos para preservar sua origem e natureza. Neste contexto as organizações públicas possuem vastas bases de dados, mas também necessitam aprimorar constantemente suas habilidades para transformá-las em informações e conhecimento que sustentem decisões mais eficazes (CALDEIRA, 2022).

Nos últimos anos, devido a esta convergência de fatores, surgiu a possibilidade de publicar dados, especialmente públicos, em formato de dados abertos. Estes representam uma abordagem para disponibilizar informações governamentais de maneira que possam ser facilmente reaproveitadas. O objetivo é promover a transparência e encorajar maior participação cívica, além de incentivar o desenvolvimento colaborativo de aplicações pela sociedade (GOV, 2023).

Apenas para ilustrar, no que diz respeito aos dados abertos acessíveis por meio do Portal dados.gov.br, encontramos uma oferta de mais de 12 mil conjuntos de dados, atendendo a diversos temas que incluem desde defesa e segurança até indústria, passando por meio ambiente, educação, saúde, entre outros (GOV, 2023).

Diversas vantagens e benefícios são associados às iniciativas de publicação de dados em formato aberto. Autores como Janssen, Charalabidis, Zuiderwijk (2012) categorizam essas vantagens em três grandes grupos: políticas e sociais, econômicas, e técnicas e operacionais.

Além disso, a disponibilização de dados em formato aberto para terceiros é vista como uma oportunidade de inovação.

Chesbrough (2003) popularizou o termo inovação aberta, argumentando que os limites entre organizações e seus ambientes estão se tornando mais instáveis. Nesse cenário, o conhecimento e a tecnologia externos a uma organização desempenham um papel crucial no processo de inovação aberta. Um ambiente propício para a inovação aberta permite que diversas organizações colaborem na criação de novos serviços e produtos, sendo a disponibilização de dados entre essas organizações um fator-chave.

Os pressupostos de dados abertos e inovação aberta convergem, de forma que os dados disponibilizados podem ser utilizados por terceiros com objetivo de promover novas funcionalidades para estes dados, com possibilidade de gerar novos produtos e serviços (RUIJER; GRIMMELIKHUIJSEN; MEIJER, 2017).

Assim, o presente trabalho teve como objetivo verificar a possibilidade da utilização de dados públicos, em formato aberto, fazendo uso das potencialidades da ciência de dados, de forma a melhorar a gestão pública, visando avançar no entendimento, na aplicação prática e na promoção de inovações colaborativas.

2. REFERENCIAL TEÓRICO

Neste tópico serão abordados os conceitos que sustentam o desenvolvimento do trabalho. Assim, serão abordados os seguintes tópicos: A administração pública; abertura de dados e finalmente análise/ciência de dados.

2.1 A administração pública

A administração pública, em suas diversas esferas, figura entre as maiores detentoras de dados, estes estão em constante movimento, seja de maneira formal ou informal, em diferentes níveis de organização. Antes, esses dados eram recebidos, coletados, processados e, por vezes, descartados sem gerar grande repercussão, mas essa realidade mudou significativamente na última década (DA SILVA CRISTÓVAM; HAHN, 2020).

Quando falamos de administração pública, estamos nos referindo ao sentido objetivo, conforme definido por Carvalho Filho (2013). Esse entendimento objetivo, prévio ao subjetivo, está diretamente ligado à preocupação e à aplicação do uso sustentável de dados, que representam a impressão digital do pensamento e comportamento da sociedade em um

determinado momento. Esses dados são essenciais para compreender como a sociedade reagirá nos momentos subsequentes.

A concepção de valor público, também conhecida como *public value*, teve seu surgimento no final dos anos 1980. Em 1995, Mark Moore lançou o livro: Gerando Valor para a Sociedade: Estratégias de Gestão Governamental, introduzindo o termo da estrutura de valor público que desde então tem sido amplamente adotado em debates sobre políticas públicas e gestão (ALFORD; O'FLYNN, 2009).

A noção de valor público é frequentemente explorada por pesquisadores em diversas áreas, como economia, marketing, contabilidade, estratégia, finanças, gestão e sistemas de informação. Diferentemente do seu uso nestes campos, o conceito de valor público tem suas raízes na administração pública (AL-HUJRAN et. al., 2015). Como um paradigma na gestão pública, o valor público enfatizou a ideia de que os gestores públicos deveriam tomar decisões de forma colaborativa, buscando desencadear resultados que beneficiassem o público em geral (WILLIAMS; SHEARER, 2011; CORDELLA; BONINA, 2012).

No entanto, surgiram desafios, pois as ideias do setor privado nem sempre se adaptavam bem aos processos e práticas dos governos. Além disso, persistia a incerteza sobre quem eram os clientes, como poderiam ser melhor atendidos e qual era a importância disso em relação à missão geral (MOORE, 2013).

O movimento global em torno da disseminação de ferramentas de Tecnologia da Informação (TICs) e dos instrumentos digitais para conservação e armazenamento de dados, com capacidades cada vez maiores, trouxe uma nova perspectiva ao Estado sobre o potencial gerencial, econômico e direcional das atividades públicas. Além disso, percebeu-se que dados podem ser transformados em informações úteis, não se limitando apenas a dados isolados (FRAZÃO, 2019). Essa perspectiva, que destaca a relevância de quem detém o maior volume de dados e informações visando melhorar os processos de tomada de decisão, em busca de maior eficácia e eficiência na Administração Pública.

2.2 Abertura de dados

Com o intuito de atender às crescentes demandas da sociedade por maior transparência, observou-se a implementação do processo de abertura de dados em diversos países, incluindo o Brasil, Estados Unidos e Reino Unido (MORELLI et al., 2017). Conforme apontado por Huijboom e Van den Broek (2011) e Van Loenen (2018), o movimento em prol

da abertura de dados também tem como foco central outros objetivos, como promover maior participação social e impulsionar a eficiência e inovação governamentais.

Os autores destacam que, em alguns países, como Estados Unidos, Reino Unido, Espanha, Dinamarca e Austrália, essa estratégia se baseia em três eixos principais, que são os seguintes: a) aumentar o controle social e incentivar a participação política; b) estimular o serviço público e a inovação de produtos; e c) reforçar a aplicação da lei.

A primeira busca capacitar os cidadãos por meio da publicação de dados governamentais, permitindo o exercício pleno de seus direitos. A segunda visa criar oportunidades de inovação impulsionadas por dados governamentais abertos, enquanto a terceira busca envolver os cidadãos e fortalecer a aplicação da lei e a segurança. A análise de Huijboom e Van den Broek (2011) revela contextos políticos e institucionais que desempenharam papel crucial na adoção da estratégia de abertura de dados

No contexto brasileiro, há uma distância considerável dessa realidade devido às marcantes desigualdades no país (PEREIRA; BRAGA; NADALIN, 2019). Fatores como a baixa qualidade na educação, elevados índices de corrupção, reduzida participação social e a instabilidade política e institucional ao longo da última década (2010-2020) demandam uma abordagem de dados abertos que fortaleça a democracia e a cidadania. Isso deve ocorrer sem prejudicar a visão empreendedora do uso dos dados governamentais, o que, por sua vez, fortalece o ecossistema dos dados abertos ao viabilizar e ampliar o uso das informações públicas.

Segundo o W3C (2009), dados abertos referem-se a dados disponibilizados em formatos pré-definidos, facilitando seu uso por terceiros em dispositivos tecnológicos integrantes das TICs. A divulgação de dados em formato aberto apresenta o potencial de contribuir para diversas iniciativas, como inovação, transparência, colaboração e participação. A transparência possibilita que terceiros acessem os dados, enquanto a participação permite que contribuam com ideias e competências. A colaboração aprimora a eficácia, estimulando a cooperação (HARRISON; SAYOGO, 2013).

De acordo com Chesbrough (2003), o compartilhamento de dados é crucial para impulsionar a inovação aberta. Ao aumentar a transparência, a colaboração e a participação entre organizações e diferentes atores da sociedade, a chance de inovação aberta se expande.

No relatório da OCDE (2019), divulgado em novembro de 2019, a Organização propôs uma aproximação mais estreita entre o governo e a população através do uso estratégico de dados. Isso possibilitaria aos Estados criarem condições para: i) aprimorar a qualidade dos serviços públicos; ii) aumentar a eficácia dos gastos públicos, mantendo

salvaguardas éticas; e iii) desenvolver um manejo seguro e eficiente de dados. Esse processo tem como objetivo não apenas preservar ou estabelecer uma zona de confiança pública, mas também proporcionar resultados efetivos na gestão.

2.3 Análise/Ciência de dados

Nas últimas décadas, testemunhamos avanços tecnológicos em um ritmo sem precedentes na história humana, especialmente no campo das Tecnologias de Informação e Comunicação (TICs). Esses progressos moldaram novos hábitos e formas de interação das pessoas com a tecnologia, resultando em uma profusão acelerada de informações. Embora tenham tornado o acesso à informação mais democrático e simplificado em diversos aspectos, essas mudanças também trouxeram desafios para o uso, mediação, organização e recuperação de informações. Destaca-se a dificuldade em distinguir as informações relevantes para as necessidades de um usuário e aquelas que são dispensáveis (MATOS; CONDURÚ; BENCHIMOL, 2022).

Diante desse cenário, surge, por volta do início dos anos 2000, uma nova disciplina denominada ciência de dados. Essa área, marcada pela interdisciplinaridade, tem como objetivo extrair valor a partir de dados, buscando conhecimento sobre eles para fornecer informações concretas que auxiliem na tomada de decisões (CAO, 2016).

O avanço tecnológico e as mudanças na percepção científica, influenciadas pelo movimento de ciência aberta (Smart *et al.*, 2019), levam à discussão sobre a abertura e compartilhamento de dados científicos. Essa iniciativa visa criar novos conhecimentos, fomentar a participação científica e possibilitar inovações. A ciência de dados, como uma área tecnológica centrada em dados digitais (especialmente dados abertos), se alinha aos princípios da ciência aberta e dos dados abertos devido à importância da reprodutibilidade, transparência de dados e à necessidade de inovação.

Conforme Moreira, Beira e Oliveira (2020), a concepção de dados, informação e conhecimento é fundamental para a comunicação. Os autores resumem que os dados são considerados a matéria-prima para a informação, que, por sua vez, é vista como a matéria-prima para o conhecimento, seguindo uma hierarquia. Desse modo, o conhecimento pode ser entendido como a percepção da realidade derivada de uma informação específica.

Os dados são descritos como informações brutas, carentes de significado antes de serem processados. Segundo Santos D'Amorim *et al.* (2020), são observações específicas ou resultados de medições que, isoladamente, não conseguem transmitir uma mensagem

compreensível. Assim, quando vistos de maneira isolada, seu real significado não é compreendido.

Nesse contexto dinâmico, em que avanços constantes e a demanda por aprimoramento contínuo são prementes, o compartilhamento de conhecimento torna-se essencial para o desenvolvimento da ciência de dados. Diante desse panorama, é relevante investigar o estado da produção científica no campo da ciência da informação relacionada à ciência de dados. Isso permitirá compreender as influências e relações entre essas duas disciplinas, além de avaliar a produção de acesso aberto relacionada ao tema. Para isso, a bibliometria se apresenta como uma ferramenta crucial na análise e avaliação da produção e comunicação científicas, possibilitando identificar avanços e tendências na ciência (MATOS; CONDURÚ; BENCHIMOL, 2022).

Atualmente, conceitos como Big Data e Data Science estão cada vez mais presentes e são frequentemente considerados fenômenos tecnológicos. Isso se deve ao fato de que cada usuário contribui significativamente para a produção massiva de dados.

De acordo com Raudemberg e Carmo (2019), "Big Data refere-se a um conjunto de dados que é tão extenso que impede a captura, armazenamento, gerenciamento e análise por meio de ferramentas computacionais tradicionais. Ele demanda abordagens inovadoras para processar grandes volumes de dados heterogêneos". Por outro lado, a Ciência de Dados, também conhecida como Data Science, é definida pela Oracle (Oracle, 2021) como "um subconjunto da inteligência artificial (IA) que se refere principalmente às áreas sobrepostas de estatísticas, métodos científicos e análise de dados - todas utilizadas para extrair significado e insights dos dados".

Portanto, ter um grande volume de dados não garante automaticamente que todos esses dados representem conhecimento. Podemos entender a ciência de dados como a extração de informações a partir da base de dados do *big data*, resultando em dados complementares que se transformam em informação significativa.

Diante da crescente relevância e atratividade da carreira de cientista de dados, mesmo sendo um termo relativamente novo, observa-se uma rápida expansão de cursos de ciência de dados em universidades no Brasil (ARA, 2023). A urgência do mercado e a elevada demanda por profissionais qualificados impulsionam essa evolução, refletida no aumento significativo do número de cursos registrados no Cadastro Nacional de Cursos e Instituições de Educação Superior (Cadastro e-MEC).

Em novembro de 2020, eram contabilizados pelo menos 36 cursos de graduação em Ciência de Dados no país, número que, em 2023, já ultrapassa os 70 cursos, incluindo aqueles

que estão por iniciar. Essa expansão representa uma resposta eficaz do sistema educacional às necessidades emergentes do mercado.

3 METODOLOGIA

Neste tópico apresenta-se uma abordagem sobre revisão bibliográfica e estão descritos os procedimentos metodológicos desenvolvidos para que o trabalho atenda ao objetivo. Assim, o trabalho é conceituado metodologicamente, e descrito o procedimento de coleta e análise de dados.

3.1 Revisão bibliográfica

A revisão bibliográfica desempenha um papel crucial na redação de um texto científico, independentemente de seu gênero, seja uma tese, uma dissertação, ou um artigo científico de revisão (EQUIPE EDITORIAL, 2020). Noronha e Ferreira (2000), ao analisarem a produção bibliográfica, destacam a importância da temporalidade nas áreas temáticas. Essa abordagem possibilita fornecer um panorama atualizado sobre um tópico específico, revelando novas ideias e métodos com maior ou menor destaque na literatura especializada.

Conforme Yin (2016), a pesquisa bibliográfica consiste em "uma incursão inicial para construir um banco de estudos qualitativos previamente realizados, a fim de auxiliar na consideração do tema, do método e das fontes de evidências para uma nova investigação". Esse tipo de pesquisa baseia-se em referências teóricas já examinadas e publicadas em diversas fontes, permitindo que o pesquisador compreenda o que já foi estudado sobre o assunto.

Quando se trata da elaboração de um artigo de revisão, a leitura aprofundada e intensa dos textos a serem utilizados, conforme observado por Motta Roth e Hendges (2010), pode oferecer diversas contribuições para identificar as relações entre as pesquisas mencionadas. Isso envolve questionamentos como: há sobreposições, contrastes, complementações? Segundo as autoras, é crucial que os autores citados estabeleçam um diálogo entre si por meio da mediação discursiva dos autores citados, explorando o quão essenciais podem ser para a contribuição ao tema em análise.

3.2 Coleta e análise de dados

Para os procedimentos de coleta de dados foram realizadas buscas em 4 (quatro) bases de dados, sendo estas: *Scopus*, *SciELO*, *Web of Science* e Banco Digital de Teses e Dissertações (BDTD).

No campo de pesquisa das bases de dados foi realizada uma busca pelas palavras, combinadas em três conjuntos, quais sejam: a) “ciência” e “aberta”; b) “dados” e “governo”; e c) “dados” e “governamentais”. Todas as três combinações na caixa do campo título. Todos os trabalhos deveriam estar escritos no idioma português, não foi colocado um corte temporal para as buscas. Os trabalhos localizados foram inseridos em uma planilha eletrônica, incluindo o título, autores, resumo, palavras-chaves e objetivo. Alguns deles não apresentavam resumo, ou não continham informações pertinentes a este trabalho, então foram eliminados.

A primeira base de dados pesquisada foi a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD – <https://bdtb.ibict.br>). A BDTD trata-se de um portal de busca que integra e dissemina os textos completos das teses e dissertações defendidas nas instituições brasileiras de ensino e pesquisa, é uma iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT que busca dar maior visibilidade da produção científica nacional e a difusão de informações de interesse científico e tecnológico para a sociedade em geral.

A segunda base de dados a ser pesquisada foi a *Scopus* (*Scopus* – <https://www.scopus.com>). Esta base de dados é um repositório de trabalhos revisados por pares, incluindo revistas científicas, livros e anais de conferências. Oferece uma visão abrangente da produção mundial de pesquisa nas áreas de ciência, tecnologia, medicina, ciências sociais e artes e humanidades.

Posteriormente foram realizadas pesquisas na base de dados *Scientific Electronic Library Online* (*SciELO* – <https://www.scielo.br>). Esta base de dados criada 1997, adotada em dezenas de países que formam a Rede *SCIELO* de coleções nacionais de periódicos.

A quarta, e última base de dados pesquisa, foi a *Web of Science* (*Web of Science* – <https://www.webofscience.com>). O *Web of Science* (anteriormente conhecido como *Web of Knowledge*) é uma base de dados com assinatura a vários bancos de dados que fornecem dados abrangentes de citações para muitas disciplinas acadêmicas diferentes.

Para análise de dados foram adotadas as premissas da análise de conteúdo conforme Bardin (2009). De acordo com essa autora, análise de conteúdo é uma técnica de investigação

que tem por finalidade a descrição objetiva e sistemática do conteúdo manifesto da comunicação. Uma das formas de operacionalizar a análise de conteúdo é realizar a categorização dos textos advindos da coleta de dados. Essa categorização pode ser realizada a priori, como foi neste trabalho, sendo as categorias construídas conforme as palavras chaves utilizadas nas buscas nas bases de dados.

4 ANÁLISE E DISCUSSÕES DOS RESULTADOS

Nesta seção serão apresentados os resultados das quatro bases de dados utilizadas para o processo de coleta de dados. Na Tabela 1 (Resultado das pesquisas realizadas) apresenta-se um resumo com o resultado das pesquisas realizadas nas bases de dados. Nesta tabela já foram desconsiderados os artigos com: resumos indisponíveis e artigos repetidos.

TABELA 1 - Resultado das pesquisas realizadas

Bases de dados	Palavras chaves usadas na busca		
	Ciência + Aberta	Dados + Governo	Dados + Governamentais
BDTD	18	49	70
SCOPUS	4	18	37
SciELO	4	6	14
WEB OF SCIENCE	0	2	0
Totais	26	75	121

Fonte: Autores do trabalho (2024)

Após este primeiro corte (artigos sem resumos e repetidos), foi realizada uma leitura em todos os resumos, para verificar quais trabalhos efetivamente tinham/tem relação com o objetivo deste trabalho. Assim, nas tabelas estão (nas próximas seções) os trabalhos que fazem alusão a possibilidade da utilização de dados governamentais, em formato aberto, para análise de dados, visando descobertas conforme conceitos da ciência de dados. Na sequência são apresentados, conforme a base de dados, os trabalhos que restaram após a leitura de cada resumo.

Na Tabela 2 apresenta-se os resultados da pesquisa realizada na base de dados BDTD. Em todas as tabelas está exposto apenas título do trabalho, por uma questão de limitação de

espaço. Embora as análises apresentadas considerem também o resumo. Para acesso aos resumos e demais dados de cada trabalho, realizar contato os autores do trabalho.

TABELA 2 – Resultados da pesquisa realizada na base de dados BDTD

Título do trabalho
ANÁLISE DE CONTAS PÚBLICAS UTILIZANDO DADOS ABERTOS GOVERNAMENTAIS: UM ESTUDO DE CASO NO GOVERNO DO ESTADO DO CEARÁ
Mineração de dados aplicados ao sistema integrado de administração financeira do governo federal - SIAFI : detecção de anomalias na emissão de notas de empenho
Utilização de banco de dados ontológicos e análise de redes sociais de cidadãos em sistemas de governo eletrônico
Precisamos encontrá-los! : etnografia dos números do Cadastro Único e dos cruzamentos de base de dados do governo federal brasileiro
Avaliação da eficiência dos gastos do governo na educação primária e secundária dos países da América Latina comparada aos países europeus utilizando análise envoltória de dados
Além das “suposições” : medindo a eficiência da gestão orçamentária das unidades do orçamento fiscal e da seguridade social do governo do Distrito Federal por meio do cálculo do Índice de Qualidade da Gestão Orçamentária e da análise envoltória de dados
Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação
Modelo de apoio à decisão para avaliar dados governamentais abertos do setor elétrico
Data science aplicada à análise criminal baseada nos dados abertos governamentais do Brasil
ANÁLISE DE CONTAS PÚBLICAS UTILIZANDO DADOS ABERTOS GOVERNAMENTAIS: UM ESTUDO DE CASO NO GOVERNO DO ESTADO DO CEARÁ
DA AQUISIÇÃO A VISUALIZAÇÃO DE DADOS: APLICAÇÕES DA CIÊNCIA DE DADOS EM SAÚDE
Ciência de dados e políticas públicas de saúde: exemplos práticos
Compreendendo e prevendo o processo legislativo via ciência de dados
Utilização da mineração de dados para identificar a evasão nos cursos EaD do Instituto Federal de Educação, Ciência e Tecnologia Farroupilha
O uso de ciência de dados e inteligência geográfica como metodologias de políticas públicas para o diagnóstico precoce de tumores

Fonte: Autores do trabalho (2024)

Conforme os dados da Tabela 1, a primeira pesquisa na base BDTD, resultou em 137 trabalhos. Após a leitura do resumo dos mesmos, resultaram em 15 trabalhos (Tabela 2). A análise dos títulos e resumos revela:

- a) cinco trabalhos estão suportados pelo uso de dados abertos, este fato reforça a relação entre ciência aberta e dados abertos;
- b) cinco trabalhos estão relacionados com questões orçamentárias e quatro com a área da saúde.

c) demais trabalhos tem foco em áreas/temas variados, tais como: educação, setor elétrico, segurança, entre outros.

Na Tabela 3 apresenta-se o resultado da pesquisa realizada na base de dados *SciELO*. Em todas as tabelas está exposto apenas título do trabalho, por uma questão de limitação de espaço. Embora as análises apresentadas considerem também o resumo. Para acesso aos resumos e demais dados de cada trabalho, realizar contato os autores do trabalho.

TABELA 3 - Pesquisa realizada na base de dados SciELO

Título
Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde?
Vigilância e privacidade, no contexto do big data e dados pessoais: análise da produção da Ciência da Informação no Brasil

Fonte: Autores do trabalho (2024)

Conforme os dados da tabela 1, a primeira pesquisa na base *SciELO*, resultou em 24 trabalhos. Após a leitura do resumo dos mesmos, resultaram em apenas 2 trabalhos (Tabela 3). Destes, um trabalho explora as potencialidades do uso do Big Data para estudos sobre a sociedade, em especial, na área da saúde. O segundo trabalho também é suportado pela tecnologia *big data*, relacionado com o tema/área de segurança.

Na Tabela 4 apresenta-se o resultado da pesquisa realizada na base de dados *Scopus*. Em todas as tabelas está exposto apenas título do trabalho, por uma questão de limitação de espaço. Embora as análises apresentadas considerem também o resumo. Para acesso aos resumos e demais dados de cada trabalho, realizar contato os autores do trabalho.

TABELA 4 - Pesquisa realizada na base de dados Scopus

Título
Dos dados ao conhecimento: Tendências da produção científica sobre big data em ciência da informação no Brasil
Desigualdades escolares e bem-estar urbano: indo além do status socioeconômico com a ciência de dados

Fonte: Autores do trabalho (2024)

Conforme os dados da Tabela 1, a primeira pesquisa na base *Scopus*, resultou em 59 trabalhos. Após a leitura do resumo dos mesmos, resultaram em apenas 2 trabalhos (Tabela 4). Destes, um trabalho explora as potencialidades do uso do *big data* para estudos sobre a sociedade. O segundo trabalho tem foco na área/tema educação.

Embora na Tabela 1 tenham sido encontrados dois trabalhos na base *Web of Science*, após a leitura do resumo verificou-se que os dois não tinham aderência ao objetivo deste trabalho.

5 CONCLUSÕES

O trabalho investigou a viabilidade do uso de dados governamentais abertos aliado as capacidades da Ciência de Dados. Através de buscas em quatro bases de dados (BDTD, *Scopus*, *SciELO* e *Web of Science*). Os resultados revelaram uma quantidade significativa de estudos relacionados à utilização de dados governamentais abertos para análise de dados, abrangendo diversas áreas como saúde, educação, segurança e gestão pública. Isso ressalta o potencial desses dados para suportar decisões, em busca de melhores resultados.

Observa-se uma diferença em números de trabalhos encontrados em cada base de dados, destacando-se a BDTD como a que apresentou o maior número de estudos relevantes para o tema. Pode-se concluir que a utilização de dados governamentais abertos para análise de dados é uma área de pesquisa em crescimento, com potencial para impactar positivamente a gestão pública e contribuir para o avanço do conhecimento científico em várias áreas.

Como limitações do estudo, podemos indicar a consulta em quatro bases de dados apenas, com relação a lacunas de pesquisa, sugerem-se para futuros trabalhos, pesquisas e análises em outras bases de dados.

6 REFERÊNCIAS

ALFORD, J.; O'FLYNN, J. **Making sense of public value: Concepts, critiques and emergent meanings.** Intl Journal of Public Administration, v. 32, n. 3-4, p. 171-191, 2009.

AL-HUJRAN, O.; AL-DEBEI, M. M.; CHATFIELD, A.; e MIGDADI, M. **The imperative of influencing citizen attitude toward e-government adoption and use.** Computers in Human Behavior. 53, 189-203, 2015.

ARA, A.; KATH, G. K. G.; ANJOS, C. P. dos.; RUSSO, C. M.; BONAT, W. H. **Ciência de Dados: uma descrição dos primeiros cursos de graduação em ciência de dados nas universidades brasileiras.** SciELO Preprints , 2023. DOI: 10.1590/SciELOPreprints.6570. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/6570> . Acesso em: 28 nov. 2023.

BARDIN, L. **Análise de Conteúdo.** 9ª edição. Lisboa: editora 70, 2009.

BISPO, M. D. S. **A impossibilidade da ciência aberta sem alteridade e pluralidade epistêmica.** Revista de Administração Contemporânea. v. 26, 2021.
<https://doi.org/10.1590/1982-7849rac2022210246>.

CADASTRO e-MEC. **CADASTRO NACIONAL DE CURSOS E INSTITUIÇÕES DE EDUCAÇÃO SUPERIOR.** Cadastro e-MEC. Disponível em: <<https://emec.mec.gov.br/>> .
Acesso em: 29 de nov de 2023.

CAO, L. **Data science: a comprehensive over - view.** ACM Computing Surveys, v. 50, n. 3, jun. 2018. Disponível em: <https://dl.acm.org/doi/10.1145/3076253>. Acesso em: 10 jan 2024.

CARVALHO FILHO, J. dos S. **Manual de direito administrativo.** 26. ed. São Paulo: Atlas, 2013.

CHESBROUGH, H. **Open Innovation The New Imperative for Creating and Profiting from Technology.** Boston, Massachusetts: Harvard business school press, 2003.

CORDELLA, A.; BONINA, C. M. **A public value perspective for ICT enabled public sector reforms: A theoretical reflection.** Government Information Quarterly, 29(4), 512-520, 2012.

DA SILVA CRISTÓVAM, J. S.; HAHN, T. M. **Administração Pública orientada por dados: Governo aberto e infraestrutura nacional de dados abertos.** Revista de Direito Administrativo e Gestão Pública, v. 6, n. 1, p. 1-24, 2020.

EQUIPE EDITORIAL. Editorial. **Interações** (Campo Grande), [S. l.], v. 21, n. 4, p. 681–684, 2020. DOI: 10.20435/inter.v21i4.3203. Disponível em:
<https://interacoesucdb.emnuvens.com.br/interacoes/article/view/3203>. Acesso em: 28 nov. 2023.

FRAZÃO, A. **Fundamentos da proteção dos dados pessoais.** In: FRAZÃO, A.; TEPEDINO, G.; OLIVA, M. D. (Coord). Lei Geral de Proteção de Dados e suas repercussões no direito brasileiro. São Paulo: Revista dos Tribunais, 2019.

GIL, A. C. **Métodos e técnicas de pesquisa social.** 5.ed. São Paulo: Atlas, 1999.

GOVERNO BRASILEIRO. **Sobre Dados Abertos.** Disponível em:
<<https://www.gov.br/governodigital/pt-br/dados-abertos>>. Acesso em: 07 dez. 2023.

HARRISON, T.; SAYOGO, D. **Open Budgets and Open Government: Beyond Disclosure in Pursuit of Transparency, Participation and Accountability.** The Proceedings of the 14th Annual International Conference on Digital Government Research. Quebec, Canadá, 2013.

HUIJBOOM, N.; VAN DEN BROEK, T. Open data: an international comparison of strategies. **European Journal of ePractice**, (12).
<http://unpan1.un.org/intradoc/groups/public/documents/UN-DPADM/UNPAN046727.pdf>, 2011.

JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. **Benefits, Adoption Barriers and Myths of Open Data and Open Government.** *Information Systems Management*, 29:258–268, 2012.

MATOS, M. T. de; CONDURÚ, M. T.; BENCHIMOL, A. C. **Interseções na produção científica da ciência da informação e ciência de dados.** *Acervo, [S. l.]*, v. 35, n. 2, p. 1–18, 2022. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1804>. Acesso em: 15 jan. 2024.

MOORE, M. H. **Creating Public Value:** strategic management in Government. London: Harvard University Press. 1995.

MOORE, M. H. **Recognizing Public Value.** Harvard University Press. Edição do Kindle, 2013. E-book.

MOREIRA, C.; BEIRA, J. C.; OLIVEIRA, M. **Um olhar dos estudantes do curso de biblioteconomia acerca do que são dados, Informações e Conhecimentos.** <http://www.uel.br/revistas/informacao>, [s. l.], p. 484 – 508, 10 jan. 2024.

MORELLI, N.; MULDER, I.; PEDERSEN, J. S.; JASKIEWICZ, T.; DE GÖTZEN, A.; ARGUILLAR, M. **Open Data as a New Commons. Empowering Citizens to Make Meaningful Use of a New Re-source.** In I. Kompatsiaris et al. (eds.), *Internet Science Lecture Notes in Computer Science*, 10673, 212-221. Springer. https://doi.org/10.1007/978-3-319-70284-1_17. 2017.

MOTTA-ROTH, D.; HENDGES, G. R. **Produção textual na universidade.** São Paulo: Parábola Editorial, 2010.

NORONHA, D. P.; FERREIRA, S. M. S. P. Revisões de literatura. In: CAMPELLO, B. S.; CONDÓN, B. V.s; KREMER, J. M. (Org.). **Fontes de informação para pesquisadores e profissionais.** Belo Horizonte: UFMG, 2000.

OECD. **The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, Paris:** OECD Publishing, 2019. Disponível em: <https://doi.org/10.1787/059814a7-en>. Acesso em: 2 mai. 2020.

ORACLE BRAZIL. In: **O que é Ciência de Dados?** [S. l.], 2 fev. 2021. Disponível em: <https://www.oracle.com/br/data-science/what-is-data-science/>. Acesso em: 10 jan. 2024.

PEREIRA, R. H. M.; BRAGA, C. K. V.; NADALIN, V. G. **Desigualdades Socioespaciais de Acesso a Oportunidades nas Cidades Brasileiras.** Instituto de Pesquisa Econômica Aplicada. https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_2535_web.pdf. 2019.

PEREIRA, A. L.; REZENDE, S. O.; MARCACINI, R.; CALDEIRA, D. M. **Ciência de Dados na Administração Pública: Desafios e Oportunidades.** *Revista da CGU, [S. l.]*, v. 14, n. 26, 2022. DOI: 10.36428/revistadacgu.v14i26.617 . Disponível em: https://revista.cgu.gov.br/Revista_da_CGU/article/view/617 . Acesso em: 28 nov. 2023.

RAUTENBERG, S.; CARMO, P. R. V. **Big data e ciência de dados: complementaridade conceitual no processo de tomada de decisão.** Brazilian Journal of Information Studies: Research Trends, [s. l.], 2019.

RUIJER, E.; GRIMMELIKHUIJSEN, S.; MEIJER, A. **Open data for democracy: Developing a theoretical framework for open data use.** Government Information Quarterly. Elsevier. 2017.

SANTOS-D'AMORIM, K.; CRUZ, R. W. dos R.; SILVA, M. L. da.; CORREIA, A. E. G. C. **Dos dados ao conhecimento: tendências da produção científica sobre big data na ciência da informação no Brasil: Encontros Bibli: Revista eletrônica de biblioteconomia e ciência da informação.** Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, [s. l.], 2020.

SMART, P.; HOLMES, S.; LETTICE, F.; PITTS, F. H.; ZWIEGELAAR, J.; SCHWARTZ, G.; EVANS, S. **Open science and open innovation in a socio-political context: knowledge production for societal impact in an age of post-truth populism.** R&D Management, v. 49, n. 3, p. 279-297, 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/radm.12377>. Acesso em: 10 set 2023.

TEIXEIRA, E. B. **A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais.** Desenvolvimento em questão, v. 1, n. 2, p. 177-201, 2003.

VAN LOENEN, B. **Towards a User-Oriented Open Data Strategy.** In B. van Loenen, G. Vancauwenberghe & J. Cromptvoets (eds.), Open Data Exposed (pp. 33-53). Springer. 2018.

W3C. **Manual dos Dados Abertos: governo. Traduzido e adaptado de <http://opendatamanual.org>.** Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf, 2009. Acesso em: 24 jan. 2024.

WILLIAMS, I.; SHEARER, H. **Appraising Public Value: Past, Present and futures.** Public Administration, 89(4), 1367–1384, 2011.

YIN, R. K. **Pesquisa qualitativa do início ao fim.** Penso Editora, 2016.

AGRADECIMENTOS

Agradecemos a Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul/AFAPERGS e Universidade Federal do Pampa/UNIPAMPA, pela concessão de bolsa para financiamento de pesquisa que possibilitou o desenvolvimento deste trabalho.