



A CAPTURA DE DADOS NO CONTEXTO DE WEBINÁRIO ASSOCIADO COM O “PROGRAMA DINHEIRO DIRETO NA ESCOLA”

Guilherme Ataíde Dias, Wagner Junqueira Araújo e Renata Lemos dos Anjos

Universidade Federal da Paraíba (UFPB), Brasil

RESUMO: A Ciência dos Dados possibilita a transdução dos dados em informação para o processo de tomada de decisão. Diversas instituições públicas e/ou privadas passam a fazer uso dos dados provenientes de investigações científicas em suas gestões, o que reforça a importância da garantia da qualidade dos dados. Em 2021 o Programa Dinheiro Direto da Escola iniciativa fomentada pelo Fundo Nacional de Desenvolvimento da Educação realizou o Webinar PPDE Ações Integradas, o qual orientou e esclareceu dúvidas sobre as melhores práticas para utilização dos recursos recebidos. As comunidades associadas com as escolas participantes do programa, foram convidadas a enviar suas dúvidas através de um questionário *online*. A pesquisa busca desvendar quais e como foram realizados os procedimentos para a garantia da qualidade dos dados provenientes da aplicação dos questionários do Webinar PPDE Ações Integradas da região nordeste. Verificou-se a importância do controle de consistência nos dados capturados. Campos cujo preenchimento foram apresentados de forma livre, deveriam ser disponibilizados para a comunidade como um conjunto de opções fechadas quando possível. Além da necessidade do desenvolvimento de um vocabulário controlado para garantir a consistência dos termos usados em todas as instâncias relacionadas com a captura de dados no contexto do PDDE.

Palavras-chave: Ciência dos dados, qualidade dos dados, PDDE.

DATA CAPTURE WITHIN THE WEBINAR CONNECTED TO THE “DIRECT CASH TO SCHOOL PROGRAM”

ABSTRACT: Data Science enables the transduction of data into information into the decision-making process. Several public and/or private institutions start to make use of data from scientific investigations in their management, which reinforces the importance of ensuring data quality. In 2021 the School Direct Money Program initiative promoted by the National Fund for Education Development carried out the Webinar PPDE Integrated Actions, which guided and clarified doubts about the best practices for the use of the resources received. The communities associated with the schools participating in the program were invited to submit their questions through an *online questionnaire*. The research seeks to uncover which and how the procedures were performed to ensure the quality of data from the application of the questionnaires of the Webinar PPDE Integrated Actions of the Northeast region. The importance of consistency control in the captured data was verified. Fields whose completion were presented freely, should be made available to the community as a set of closed options when possible. In addition to the need to develop a controlled vocabulary to ensure consistency of the terms used in all instances related to data capture in the context of the PDDE.

Keywords: Data science, Data quality, PDDE.

Correspondência para (correspondence to:) guilhermeataide@gmail.com

INTRODUÇÃO

Em meio a Ciência dos Dados, algumas discussões acerca do acesso aberto às investigações científicas despontam e vem ganhando destaque em toda a comunidade acadêmico-científica global. Em sintonia, o acesso aberto aos dados que subsidiam

essas pesquisas também vêm ganhando protagonismo quando, esse compromisso passa a ser reconhecido pela comunidade como uma parte essencial das boas práticas de pesquisa (SILVA *et al.*, 2018; PICCOLO *et al.*, 2021).

Para Brase e Farquhar (2011), os dados

como insumos de informação, são considerados patrimônio da ciência universal e um pilar primordial para o seu avanço. Dessa forma, o acesso aberto aos mesmos torna-se um imperativo global, em que os pesquisadores trabalham em cooperação, criam, compartilham e acessam dados globalmente.

Ainda em amparo ao acesso aberto aos dados de pesquisa, Tenopir *et al.* (2015) alegam que esse compromisso em disponibilizar os dados de pesquisa de forma aberta, traz benefícios para o pesquisador e/ou grupo de pesquisa, instituições de ensino – e outras – além de entidades financiadoras. A exemplo, possibilitando uma maior visibilidade e prestígio da produção acadêmico-científica; fundamenta o investimento ao mesmo tempo em que evita duplicação, e consequentemente controle das despesas, além de promover o avanço da ciência beneficiando não apenas àqueles sujeitos integrantes da esfera acadêmico-científica, mas a comunidade universal.

Apesar de todo um movimento global a favor da ascensão do acesso aberto aos dados, é válido ressaltar que o simples compartilhamento de dados não é o suficiente. Em virtude dessa realidade, alguns pesquisadores preocupam-se além do compartilhamento de dados, visto que, o mesmo não garante a compreensão acerca dos dados de pesquisa, bem como, o seu efetivo uso e reúso.

Neste ínterim, surgem iniciativas como a dos Princípios FAIR, que por meio de diretrizes objetivam conjuntos de dados encontráveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e reutilizáveis (*Reusable*). A adoção desses princípios vai além do mero compartilhamento de dados de pesquisa. Os princípios FAIR, permitem que os dados sejam compreendidos quando recomendam, por exemplo, que os dados sejam descritos com metadados avançados fazendo uso de linguagem formal, acessível e compartilhada, além do uso de

vocabulários controlados (WILKINSON *et al.*, 2016).

Percebe-se que, os Princípios FAIR preocupam-se e devem ser considerados, em sua maioria, no momento da publicação dos dados. Mas que sua adoção não corresponde e/ou não garante a qualidade dos dados que serão compartilhados.

Desse modo, quanto a qualidade dos dados, o W3C, apresenta em um documento intitulado “Boas Práticas para Dados na Web”¹ recomendações - 35 Boas Práticas (BP) - relacionadas à publicação e ao uso de dados na Web. A BP 6: fornecer informações de qualidade de dados, afirma que, a qualidade dos dados pode afetar diretamente o uso dos dados para aplicações específicas e/ou aplicações que diferem do propósito para qual os dados foram criados e/ou coletados. Além de recomendar o registro de todos os processos pelos quais os dados passaram.

Para Sant’Ana (2016) a qualidade dos dados deve ser considerada em todas as fases do ciclo de vida dos dados. Na fase da coleta devem ser considerados elementos como a procedência, mecanismos de coleta e garantias de integridade física e lógica, afirmando que, a confiabilidade dos dados é necessária para que sejam úteis. Na fase do armazenamento, as definições para tal são fundamentais para garantir a integridade física (em termos de capacidade) e lógica (acessos e/ou alterações indevidas). A qualidade dos dados na fase de recuperação deve ter especial atenção a interação do usuário, fazendo com que o mesmo compreenda os dados recuperados. Por fim, na fase de descarte, a qualidade deve ser considerada no momento do registro detalhado sobre a eliminação.

Neste ínterim, a Ciência dos Dados desenvolve-se por meio de uma ampla valorização, exploração, análise e uso intensivo de grandes volumes de dados de pesquisa. Essa nova forma de fazer ciência

também possibilita a transdução dos dados em informação para o processo de tomada de decisão. Nesse sentido, diversas instituições públicas e/ou privadas (de ensino, jurídica, econômica, política, entre outras) passam a fazer uso dos dados provenientes de investigações científicas em suas gestões, reforçando assim, a importância da garantia da qualidade dos dados (BELL *et al.*, 2009; RAUTENBERG & CARMO, 2019).

Assim posto, o presente estudo discute os desafios encontrados como foram abordados, visando a garantia da qualidade dos dados, obtidos por meio de questionários *online*, submetidos pelos integrantes das escolas participantes da iniciativa fomentada pelo Fundo Nacional de Desenvolvimento da Educação (FNDE) intitulada de Programa Dinheiro Direto na Escola (PDDE), durante a execução do Webinário *PPDE Ações Integradas*, realizado em agosto de 2021 com as comunidades associadas às escolas fundamentais participantes. O Webinário em foco, objetivou a orientação e o esclarecimento de dúvidas sobre a melhores práticas para a utilização dos recursos do PDDE, possibilitando assim, um melhor uso dos recursos públicos. O Webinário foi conduzido no âmbito do grupo de monitoramento do Centro Colaborador de Apoio ao Monitoramento e à Gestão de Programas Educacionais do Nordeste (FNDE/CECAMPE/NE). É pertinente mencionar que os dados após a sua coleta, higienização e processos de anonimização serão disponibilizados em um repositório aberto de dados, de modo que as dúvidas dos participantes do PDDE possam ser mais facilmente disseminadas por toda a comunidade escolar e pelos pesquisadores (especialmente na área da educação).

O PDDE foi criado em 1995 pelo FNDE e objetiva prestar assistência financeira para as escolas públicas de educação básica das redes estaduais, municipais e distrital, como também, para as escolas privadas de

ensino especial que possuam alunos matriculados na educação básica, a fim de contribuir para manutenção e melhoria da infraestrutura física e pedagógica, com consequente elevação do desempenho escolar. Além de também fortalecer a participação social e a autogestão escolar.

Diante do exposto, a pesquisa busca desvendar quais e como foram realizados os procedimentos para a garantia da qualidade dos dados provenientes dos questionários aplicados no Webinário *PPDE Ações Integradas* do Fundo Nacional de Desenvolvimento da Educação da região nordeste?

METODOLOGIA

Em agosto de 2021, foi realizado o Webinário *PPDE Ações Integradas* do FNDE que objetivou a orientação e esclarecimento de dúvidas sobre a melhor maneira de utilização dos recursos do PDDE, o qual contou com a presença de aproximadamente 14 (quatorze) mil pessoas. As comunidades associadas com as escolas participantes do programa, foram convidadas a enviar suas dúvidas por meio de um questionário *online*.

O questionário era composto por 7 (sete) questões referentes ao perfil dos integrantes, e três questões abertas referentes às dúvidas dos mesmos quanto ao PDDE.

Observou-se, após a coleta dos dados, conforme a análise ia sendo realizada, alguns pontos que iam de encontro a garantia de qualidade de dados. Os quais serão relatados a seguir, bem como, a busca para solucioná-los.

RESULTADOS E DISCUSSÃO

A primeira e a terceira questão pediam o preenchimento do atributo nome e *e-mail* dos participantes respectivamente, os quais foram preenchidos de maneiras diversas. Alguns preencheram apenas com o nome próprio, outros com nome completo, alguns as primeiras letras dos nomes maiúsculas e as outras minúsculas, e outros

com o nome todo em maiúsculo. Como solução desse problema, recomenda-se a prévia instrução de como deveria se dar o preenchimento (Ex.: Nome Sobrenome / email@emminusculo.com).

A segunda questão dizia respeito do atributo função que cada participante exercia na escola, a qual foi preenchida com diversos termos. Observou-se que,

vários participantes possuem a mesma função, mas por tratar-se de uma questão aberta, foram utilizadas terminologias distintas para uma mesma função, além da identificação de erros ortográficos. O Quadro 1, apresentado a seguir, exemplifica algumas das situações encontradas.

QUADRO 1: Terminologias Utilizadas no Atributo Função

FUNÇÃO	TERMINOLOGIAS UTILIZADAS
DIRETOR(A) DE ESCOLA	"Direror"; "Diretor"; "Diretor Adjunta"; "Diretor adjunto"; "Diretor Adjunto"; "Diretor Adjunto de escola"; "Diretor administrativo"; "Diretor Administrativo"; "Diretor Administrativo da Educação"; "Diretor de Cultura e Eventos"; "Diretor de Departamento de Gestão"; "Diretor de ensino"; "Diretor de Ensino "; "Diretor de escola Adjunto"; "Diretor de escola(a)"; "Diretor de projetos especiais"; "Diretor Escolar"; "Diretor Escolar Adjunto "; "Diretor Financeiro (Tesoureiro)"; "Diretor geral da Secretaria "; "Diretor na SEDUC "; "Diretor Territorial de Educação"; "Diretora"; "Diretora da Direc"; "Diretora adjunta"; "Diretora Adjunta"; "Diretora Adjunta escolar"; "Diretora administrativa e financeira"; "Diretora da Semed"; "Diretora da área de prestação de contas"; "Diretora de Ensino"; "Diretora do Departamento de Educação Básica"; "Diretora Financeira"; "Diretora Interina "; "Diretora NTE"; "Diretora Núcleo Territorial de Educação"; "Diretora Técnico Pedagógica da SEDUC Campina Grande - Paraíba"; "Dirigente Municipal de Educação”.
TÉCNICO(A)	"Técnica em Normatização"; "Técnico Administrativo do GPORSEFEC "; "Técnico PDDE"; "Técnica de normatização”; "Técnica pedagógica"; "Técnica pedagógica de diretoria regional de ensino da rede estadual"; "Técnico"; "Técnico da secretaria que trabalha com os PDDEs"; "Técnico da SEDUC"; "Técnico de Direc e responsável pelo Educação Conectada na Regionalização"; "Técnico de Normatização"; "Técnico Financeiro"; "Técnico pedagógica"; "Técnico(a)".

Conforme exposto no Quadro 1, nota-se o uso de diversas terminologias para uma mesma função. Esse problema foi solucionado, no processo de higienização dos dados, quando se criou categorias gerais para as funções relatadas (a exemplo: Diretor(a) de escola, Técnico(a); Conselheiro(a)) visando a garantia da qualidade. Ressalta-se que tal problema poderia ser evitado se a questão fosse de caráter fechado com as opções de respostas previamente determinadas.

A quarta, quinta e sexta questão correspondiam ao atributo faixa etária, escolaridade e tempo de envolvimento com o PDDE respectivamente, sendo fechada,

com as possíveis opções de respostas. O que já atribui uma qualidade aos dados e facilita o processo de análise, extração e visualização dos dados.

A sétima e oitava questão correspondiam a localidade dos participantes. De início foi questionada a Unidade Federativa – UF e posteriormente o município. Não foram encontradas divergências significativas quanto ao preenchimento destes campos.

A nona questão abordava se ainda havia ou não dúvida em relação ao tema apresentado. Configurando-se como uma questão fechada, com as possíveis opções de respostas, não houveram problemas

quanto à garantia da qualidade.

A décima questão, referente à quais seriam essas dúvidas, foi de caráter fechado, mas com múltiplas escolhas, ou seja, os participantes podiam selecionar quantas opções desejassem. O que também, previamente, já contribuiu para uma melhor garantida qualidade dos dados.

Por fim, a décima primeira questão, solicitava que os participantes detalhassem as suas respectivas dúvidas. Por ser uma questão de caráter aberto, foi necessário na análise dos dados, visando a sua qualidade, a criação de categorias e a classificação das dúvidas dentro dessas.

CONCLUSÕES

Verificou-se a importância em ter-se um controle de consistência nos dados capturados pelos questionários. Campos cuja forma de preenchimento foram apresentadas de forma livre, deveriam ser (nesse caso concreto) disponibilizados para a comunidade de usuários como um conjunto de opções fechadas sempre que possível. Estas opções poderiam ser selecionadas a partir da escolha dos usuários em uma lista com termos pré-selecionados. Um outro fato que se tornou evidente, foi a necessidade do desenvolvimento de um vocabulário controlado para garantir a consistência dos termos usados em todas as instâncias relacionadas com a captura de dados no contexto do PDDE. Destaca-se também que os dados a serem publicados no repositório de dados aberto da iniciativa (correntemente em desenvolvimento), deverão atender da melhor forma possível a diretrizes associadas aos Princípios FAIR e às Boas Práticas para Dados na Web desenvolvidas pelo W3C.

REFERÊNCIAS

BELL, G., HEY, T., & SZALAY, A. Beyond the data deluge. *Science*, v.323, n.5919, p.1297- 1298. 2019. <http://dx.doi.org/10.1126/science.1170411>

BRASE, J., & FARQUHAR, A. Access to research data. *D-Lib Magazine*, v.17, n.1/2. 2011. <http://doi.org/10.1045/january2011-brase>

PICCOLO, D. M. *et. al.* (). Qualidade de dados em gestão de dados de pesquisa: um estudo bibliométrico. *Em Questão*,28(1), 159–184. 2021. <https://doi.org/10.19132/1808-5245281.159-184>

RAUTENBERG, S., & CARMO, P. R. V. do. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. *Brazilian Journal of Information Science: Research Trends*, v.13. n.1, 56–67. 2019. <https://doi.org/10.36311/1981-1640.2019.v13n1.06.p56>

SANT'ANA, R. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação&Informação*, v.21. n-2, p.116–142. 2016. <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>

SILVA, L. C., SANTAREM SEGUNDO, J. E., & SILVA, M. F. Princípios FAIR e melhores práticas do Linked Data na publicação de dados de pesquisa. *Informação & Tecnologia (ITEC)*, v.5. n.2, p.81-103. 2018. <https://doi.org/10.22478/ufpb.2358-3908.2018v5n2.44812>

TENOPIR, C., *et. al.* Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoSOne*, v.10, n.8, p.1-24. 2015. <https://doi.org/10.1371/journal.pone.0134826>

WILKINSON, M. *et al.* () The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, v.3, n.60018, p.1-9. 2016. <http://doi.org/10.1038/sdata.2016.18>

NOTAS

¹ Disponível em: <https://w3c.br/traducoes/DWBP-pt-br/#DataQuality>