

## Os benefícios dos artigos de dados<sup>1</sup>

Vishwas Chavan

<https://orcid.org/0000-0002-6778-5282>

Lyubomir Penev

<https://orcid.org/0000-0002-2186-5033>

*Um artigo de dados pode ser entendido como um documento científico de uma publicação com metadados pesquisáveis, que descreve um determinado conjunto de dados acessível online, ou um grupo de conjuntos de dados, publicado de acordo com o padrão das práticas científicas.*



Fonte: Banco de imagens Pexels (2021).

Um artigo de dados é uma publicação em periódico científico, cujo objetivo principal é de descrever os dados, em vez de relatar uma investigação. Como tal, contém fatos sobre os dados, não hipóteses e argumentos em apoio a essas hipóteses baseadas em dados, como encontrado em um artigo de pesquisa convencional. Seus objetivos são três: 1. fornecer uma publicação em periódico citável, que traga crédito científico aos editores de dados; 2. descrever os dados de uma maneira

<sup>1</sup> Trecho do original intitulado: CHAVAN, V.; PENEV, L. The data paper: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, [s. l.], v. 12, n. S2, Dec. 2011. DOI: <https://doi.org/10.1186/1471-2105-12-S15-S2>.

Obs.: autorização de versão da nota em português enviada por meio do perfil de Vishwas Chavan no Researchgate - <https://www.researchgate.net/messages/63697770>



legível, estruturada de forma que seja possível a compreensão de pessoas e/ou máquinas; e 3. chamar a atenção da comunidade científica para a existência daqueles dados .

A descrição deve incluir vários elementos importantes chamados de metadados ou "descrição dos dados" que documentam, por exemplo, o conjunto de dados que foi coletado, os rótulos, etiquetas e/ou táxons relacionados, a faixa espacial e temporal e a cobertura regional dos registros de dados, informações provenientes sobre quem coletou e quem possui os dados, detalhes sobre qual *software* foi utilizado para criar os dados ou qual pode ser usado para visualizar os dados, e assim por diante. Mais detalhes podem ser vistos em Chavan e Penev (2011), artigo que apresenta uma tabela com mais de 80 metadados para descrição de dados, disponível [aqui ->](#).

Uma característica importante das publicações que contém dados é que elas devem estar sempre vinculadas aos conjuntos de dados publicados que descrevem, por meio de uma URL, com um identificador de objeto digital, DOI, publicado no próprio artigo. Por outro lado, os metadados que descrevem o conjunto de dados mantidos nos arquivos de dados devem incluir os detalhes bibliográficos, incluindo um DOI, do documento de dados uma vez que é publicado.

Muitos argumentariam que um documento de dados não é, de forma alguma, um novo conceito. A Ecological Society of America publica documentos de dados em Ecological Archives desde 2000. Earth System Science Data, CMB data papers, BMC Data Notes e International Journal of Robotics Research são algumas instâncias esporádicas de editores de dados. No entanto, um mecanismo padronizado e ferramentas de *software* associadas para gerar manuscritos de dados em artigos a partir de metadados enriquecidos que descrevem um recurso de dados ainda não estão em vigor.

Acredita-se que, se implementados na letra e no espírito, as publicações de dados podem abordar a questão do reconhecimento, um incentivo aos editores de dados por seus esforços na criação de metadados ricos de um conjunto de dados. Os

editores de dados serão creditados por meio de: (a) registro de prioridade e autoria em uma publicação acadêmica convencional em qualquer periódico adequado; (b) indexação e citação de publicações de dados da mesma forma que qualquer artigo de pesquisa, o que traz benefícios para os autores em reconhecimento e construção de carreira; (c) a capacidade de rastrear o uso e as citações de dados publicados; e (d) metadados publicados como um *paper* de dados sendo armazenados e arquivados de várias maneiras, fornecendo uma descrição persistente do recurso de dados correspondente ao longo do tempo (PENEV *et al.*, 2009; COSTELLO, 2009). Além disso, a publicação de dados permite uma divisão de trabalho em que aqueles que possuem os recursos e habilidades podem realizar os experimentos e observações necessárias para coletar conjuntos de dados potencialmente interessantes e gerenciar, curar, descobrir e publicar esses conjuntos de dados, de modo que muitas partes, cada uma com uma experiência única e capacidade de analisar os dados, pode fazer uso deles como quiserem (REES, 2010).

Os dados produzidos são coletados às custas dos esforços de pessoas e instituições, e geralmente, custeados pela sociedade. Por isso devem ser publicados, citados, usados e reutilizados separadamente ou confrontados com outros dados. Conjuntos de dados, metadados e respectivos editores de dados estão interligados para agilizar e estender mutuamente a divulgação, em benefício dos autores e da sociedade. A descoberta aumentada e direta de recursos de dados evitaria a duplicação de esforços na coleta de dados, por exemplo, das mesmas áreas ao mesmo tempo por diferentes grupos de pesquisa. Em contraste, isso abriria uma janela de colaboração entre grupos de pesquisa e entre editores de dados. Portanto, a descoberta de recursos de dados também evitará o uso indevido potencial, pois trará clareza no que diz respeito à propriedade e custódia dos dados, pois a descoberta eficiente de recursos de dados sempre trará vantagens para pesquisadores e editores de dados.

## Referências

CHAVAN, V.; PENEV, L. The data paper: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, [s. l.], v. 12, n. S2, Dec. 2011. DOI: <https://doi.org/10.1186/1471-2105-12-S15-S2>

COSTELLO, M. J. Motivating online publication of data. **BioScience**, [s. l.], v. 59, n. 5, p. 418–427, 2009. DOI: <https://doi.org/10.1525/bio.2009.59.5.9>.

PENEV, L. et al. Publication and dissemination of datasets in taxonomy: ZooKeys working example. **ZooKeys**, [s. l.], v. 11, p. 1-8, may 2009. DOI: <https://doi.org/10.3897/zookeys.11.210>

REES, J. **Recommendations for independent scholarly publication of data sets**. [s. l.]: Creative Commons Working Paper, 2010. Disponível em: <http://neurocommons.org/report/data-publication.pdf>. Acesso em: 22 may 2021.

## Dados biográficos dos autores



\* Vishwas Chavan atua como especialista em gestão do conhecimento. É reconhecido como autoridade na área de democracia da informação para o desenvolvimento sustentável. Durante 2007-2014, ele trabalhou no Global Biodiversity Information Facility como o oficial sênior no Global Biodiversity Information Facility, Universitetsparken, Copenhagen, Dinamarca.

<https://www.researchgate.net/profile/Vishwas-Chavan-2>



\*\* Lyubomir Penev atua no Instituto de Pesquisa em Biodiversidade e Ecossistemas, Academia de Ciências da Bulgária e Editores Pensoft, 13a Geo Milev Street, 1111 Sofia, Bulgária

Twitter: @LyuboPenev

<https://www.researchgate.net/profile/Lyubomir-Penev>

## Como citar

CHAVAN, V.; PENEV, L. Os benefícios dos artigos de dados. **Ciência da Informação Express**, [S. l.], v. 2, n. 6, p. 1-4, 28 jun. 2021.