



40º Simpósio Brasileiro de Bancos de Dados
40th Brazilian Symposium on Databases



SBBB
2 0 2 5



Anais do VII Dataset Showcase Workshop
(DSW 2025)

Proceedings of VII Dataset Showcase Workshop
(DSW 2025)

De 29 de setembro a 2 de outubro de 2025
September 29th to October 2nd, 2025





Organização do SBBD 2025

SBBD 2025 Organization

Program Chair

Carlos Eduardo Santos Pires (UFCG, Brasil)

Short Papers Chair

Elaine Parros M. Souza (USP, Brasil)

Demos and Applications Chair

Luiz Celso Gomes Jr (UTFPR, Brasil)

WTDBD Chair

Maria Camila Nardini Barioni (UFU, Brasil)

CTDBD Chair

Valeria Cesário Times (UFPE, Brasil)

Short Courses Chair

Dênio Duarte (UFFS, Brasil)

Tutorials Chair

Eduardo Cunha de Almeida (UFPR, Brasil)

Workshops Chair

Leonardo Andrade Ribeiro (UFG, Brasil)

WTAG Chair

Alessandreia Marta de Oliveira (UFJF, Brasil)

Proceedings Chair

Michele A. Brandão (UFMG, Brasil)

Marathon Chair

Arlino H. M. de Araújo (UFPI, Brasil)



Organização Geral *General Organization*

Chairs

Ticiania L. Coelho da Silva (UFC, Brasil)

Regis Pires Magalhães (UFC, Brasil)

Lívia Almada Cruz (UFC, Brasil)

José Antônio Fernandes de Macêdo (UFC, Brasil)



Os Anais do VII Dataset Showcase Workshop (DSW 2025) trazem os artigos selecionados e apresentados na edição do evento realizada de 29 de setembro a 2 de outubro de 2025, na cidade de Fortaleza-CE, em conjunto com o Simpósio Brasileiro de Bancos de Dados (SBBD 2025). Nesta edição, os Anais incluem oito artigos. Esses trabalhos foram selecionados por meio de um processo de revisão por pares do tipo peer-review, a partir dos 14 artigos submetidos. Este volume foi organizado por Carina F. Dorneles (UFSC), Michele A. Brandão (UFMG) e Mirella M. Moro (UFMG), coordenadoras do DSW 2025.

A principal motivação para a realização do DSW é que a publicação e a disponibilização de conjuntos de dados (abertos ou não) têm se tornado altamente relevantes devido à grande atenção dada por vários segmentos, como a mídia, indústria, academia e governo. A tarefa de disponibilização de dados é importante por inúmeros motivos, que vão desde sua reutilização em aplicativos digitais desenvolvidos pela sociedade, até a possibilidade de reprodução de experimentos desenvolvidos pela comunidade científica. Portanto, no contexto da comunidade brasileira de bancos de dados, tornar conjuntos de dados disponíveis é intrinsecamente importante devido ao fomento para novas questões de pesquisa e desenvolvimento que tal tarefa pode atingir.

Apesar de existirem diferentes conjuntos de dados disponíveis na Web, as tarefas de criar, coletar, tratar e disponibilizar dados publicamente continuam apresentando muitos desafios técnicos. Na verdade, a simples tarefa de encontrar uma boa fonte de dados e extraí-los podem ser atividades tecnicamente complexas. Talvez, mais complexo ainda, seja tratar os dados de modo a apresentá-los através de uma representação coerente e cujas informações obtidas sejam úteis e semanticamente corretas. Da mesma forma, a coleta e o tratamento dos dados para o relato de experimentos científicos para uma comunidade de banco de dados pode ser considerada uma tarefa exaustiva, trabalhosa e muito desafiadora — por outro lado, se o conjunto de dados estiver bem organizado, com uma metodologia de execução bem elaborada, a implementação e comparação de algoritmos pode ser facilitada.



Por toda essa importância dos conjuntos de dados para a realização de diferentes pesquisas, agradecemos a todas as pessoas autoras por submeterem seus trabalhos ao DSW.



The Proceedings of the VII Dataset Showcase Workshop (DSW 2025) bring the papers selected and presented at the event held from September 29 to October 2, 2025, in Fortaleza, Ceará, in conjunction with the Brazilian Database Symposium (SBBD 2025). This edition of the Proceedings includes eight papers. These papers were selected through a peer-review process from the 14 submitted papers. This volume was edited by Carina F. Dorneles (UFSC), Michele A. Brandão (UFMG), and Mirella M. Moro (UFMG), coordinators of DSW 2025.

The main motivation for holding the DSW is that the publication and availability of datasets (open or otherwise) have become highly relevant due to the significant attention given to them by various segments, such as the media, industry, academia, and government. Making data available is important for numerous reasons, ranging from its reuse in digital applications developed by society to the possibility of replicating experiments developed by the scientific community. Therefore, in the context of the Brazilian database community, making datasets available is intrinsically important due to the potential for fostering new research and development questions such a task can address.

Despite the diverse data sets available on the web, the tasks of creating, collecting, processing, and making data publicly available continue to present many technical challenges. Indeed, the simple task of finding a good data source and extracting it can be technically complex. Perhaps even more complex is processing the data to present it through a coherent representation whose information is useful and semantically correct. Similarly, collecting and processing data for reporting scientific experiments to a database community can be considered an exhaustive, laborious, and very challenging task—on the other hand, if the data set is well-organized, with a well-designed execution methodology, implementing and comparing algorithms can be facilitated.

Given the importance of datasets for conducting research projects, we would like to thank all authors for submitting their work to DSW.



Mirella M. Moro
(UFMG)

Michele A. Brandão
(UFMG)

Carina F. Dorneles
(UFSC)



Program Committee

- *Alessandreia Oliveira, UFJF*
- *André Santanchè, UNICAMP*
- *Clodoveu Davis, UFMG*
- *Daniel Dalip, CEFET-MG*
- *Daniel Lichtnow, UFSM*
- *Danilo Seufitelli, CEFET-MG*
- *Eric Araújo, UFLA*
- *Giseli Lopes, UFRJ*
- *Helena Ribeiro, UCS*
- *Isabela Gasparini, UDESC*
- *João Paulo Clarindo, USP*
- *Julio C. S. Reis, UFV*
- *Lucio Fernandes Dutra Santos, IFNMG*
- *Luiz Dias, UFF*
- *Mirela Cazzolato, USP*
- *Ronaldo Mello, UFSC*
- *Vanessa Lago Machado, IFSul*
- *Veronica Santos, IBGE/ENCE*
- *Victor Stroele, UFJF*

DSW

Conjunto de notícias para avaliação de investimentos em regiões do estado de São Paulo	1 dsw:01
<i>Gabriel Melo, João Victor Cosme Neres de Sousa, Willian Dener de Oliveira, Lucas Malta Mingardo, Carlos Eduardo Torres Freire, Agma Juci Machado Traina, Caetano Traina Jr.</i>	
EggQuality-UFRPE: A Chicken Egg Dataset for Multidisciplinary Studies in Animal Science and Computing	12 dsw:02
<i>Katariny Lima de Abreu, Andréa Silva Marques de Souza, Matheus Rocha do Carmo, Lilian Francisco de Souza, Carlos Bôa-Viagem Rabello, Denis Salvadeo, João Silva do Monte Lima, Kellyton Brito</i>	
Construindo um Dataset Georreferenciado para a Educação Superior no Brasil ..	22 dsw:03
<i>Maria Ramos, Diêgo Correia, Rafael Luciano Silva, Fabio J. Coutinho</i>	
DataCoin: Dataset Sazonal para Estratégias Quantitativas com Bitcoin	34 dsw:04
<i>Arlino Magalhaes, José Jeovane R. Cordeiro, Guilherme A. Avelino</i>	
DIP-BR: An Open and Network-Based Dataset of Brazilian Patents	46 dsw:05
<i>Iuri Veras, Pablo Vasconcelos, Danilo B Seufitelli, Daniel Hasan Dalip, Fabricio Campos</i>	
LegisPL-BR - Dataset de projetos de leis brasileiros	58 dsw:06
<i>Juan Morysson Viana Marciano, Arlino Magalhaes, Vinicius Machado Ponte</i>	
Medical Dialogue Audio Transcription: Dataset and Benchmarking of ASR Models	71 dsw:07
<i>Aline Elí Gassenn, Luís Gustavo Modelli de Andrade, Douglas Teodoro, José F. Rodrigues-Jr</i>	
DataCrimeBR: Construção de um Dataset de Crimes Reportados em Tweets no Brasil	83 dsw:08
<i>Miguel A. R. e Silva, Philipe de Freitas Melo, Thais R. M. Braga Silva</i>	