

DATA NOTE

Open Access



A comprehensive integrated dataset on Brazilian health facilities: from 2005 to 2021

Rebecca Salles^{1,2*†}, Victor de Paula Dornellas Ribeiro^{1,8†}, Raquel Gritz^{1,3†}, Carlos Leonardo Souza Cardoso^{1,3}, Balthazar Paixão^{1,2}, Lucas Carraro¹, Sérgio Ricardo de Borba Cruz¹, Vinicius Kreischer^{1,8}, Carmen Lúcia Corrêa Bonifácio^{1,3}, Raphael de Freitas Saldanha¹, Gabriel Souto^{1,2}, Matheus Miloski^{1,6}, Leandro Zirondi^{1,2}, Gizelton Pereira Alencar^{1,7}, Ariane Camilo Pinheiro Alves¹, Nelson Niero Neto¹, Jefferson da Costa Lima¹, Marcel Pedroso¹, Ronaldo Fernandes Santos Alves^{4†}, Patricia de Moraes Mello Boccolini^{5†} and Cristiano Siqueira Boccolini^{4†}

Abstract

Objectives The National Registry of Healthcare Facilities is a system with the registry of every healthcare facility in Brazil with information on the capacity building and healthcare workforce regarding its public or private nature. Despite being publicly available, it can only be accessed in separated disjoint tables, with different primary units of analysis. The objective is to offer an interoperable dataset containing monthly data from 2005 to 2021 with information on healthcare facilities, including their physical and human resources, services and teams, enriched with municipal information.

Data description Database with historical data and geographic information for each health facility in Brazil. It is composed by 5 distinct tables, organized according to combinations of time, space, and types of resources, services and teams. This database opens up a range of possibilities for research topics, from case studies in a single health facility and period, analysis of a group of health facilities with characteristics of interest, to a broader study using the entire dataset and aggregated data by municipality. Furthermore, the fact that there is a row for each health facility/month/year facilitates the integration with other datasets from the Brazilian healthcare system. In addition to being a potential object of study in the health area, the dataset is also convenient in data science, especially for studies focused on time series.

Keywords Health facilities, Data integration, Healthcare system

[†]Rebecca Salles, Victor de Paula Dornellas Ribeiro, Raquel Gritz, Ronaldo Fernandes Santos Alves, Patricia de Moraes Mello Boccolini and Cristiano Siqueira Boccolini contributed equally to this work.

*Correspondence:

Rebecca Salles
rebecca.salles@icict.fiocruz.br

¹ Platform of Data Science applied to Health (PCDaS), Institute of Scientific and Technological Communication and Information on Health (ICICT), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil

² Federal Center for Technological Education of Rio de Janeiro, CEFET/RJ, Rio de Janeiro, Brazil

³ Data Extreme Lab, National Laboratory for Scientific Computing, DEXL Lab/LNCC, Petrópolis, Brazil

⁴ Institute of Scientific and Technological Communication and Information on Health (ICICT), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil

⁵ Center for Information, Public Policies and Social Inclusion, NIPPIS, FMP/UNIFASE, Petrópolis, Brazil

⁶ Federal University of Rio de Janeiro, UFRJ, Rio de Janeiro, Brazil

⁷ School of Public Health, University of São Paulo, FSP/USP, São Paulo, Brazil

⁸ National Laboratory for Scientific Computing, LNCC, Petrópolis, Brazil



Table 1 Overview of data files/datasets

Label	Name of data file/dataset	File type (file extension)	Data repository and identifier	References
Dataset 1	CNES Beds (50M records, 451 attributes)	Compressed files (.zip)	Synapse: https://doi.org/10.7303/syn32209177.1	[5]
Dataset 2	CNES Equipments (50M records, 562 attributes)	Compressed files (.zip)	Synapse: https://doi.org/10.7303/syn32208612.1	[6]
Dataset 3	CNES Professionals (49M records, 656 attributes)	Compressed files (.zip)	Synapse: https://doi.org/10.7303/syn32210090.1	[7]
Dataset 4	CNES Services (49M records, 849 attributes)	Compressed files (.zip)	Synapse: https://doi.org/10.7303/syn32211006.1	[8]
Dataset 5	CNES Teams (55M records, 694 attributes)	Compressed files (.zip)	Synapse: https://doi.org/10.7303/syn32207908.1	[9]
Data file 1	cnes_beds_dictionary	Delimited text (.csv)	Synapse: https://doi.org/10.7303/syn32167629.1	[10]
Data file 2	cnes equipments_dictionary	Delimited text (.csv)	Synapse: https://doi.org/10.7303/syn32167630.1	[11]
Data file 3	cnes_professionals_dictionary	Delimited text (.csv)	Synapse: https://doi.org/10.7303/syn32167631.1	[12]
Data file 4	cnes_services_dictionary	Delimited text (.csv)	Synapse: https://doi.org/10.7303/syn32167632.1	[13]
Data file 5	cnes_team_dictionary	Delimited text (.csv)	Synapse: https://doi.org/10.7303/syn32167633.1	[14]
Data file 6	CNES_ETL_report	Portable document (.pdf)	Synapse: https://doi.org/10.7303/syn34388050.2	[15]
Data file 7	ETL_example_CNES_EP	Jupyter notebook (.ipynb)	Synapse: https://doi.org/10.7303/syn34388051.1	[16]

Objective

Approximately half the world's population lacks access to healthcare services, medicines, and vaccination [1]. Brazil includes universal health coverage in its Constitution, which is operated through the Unified National Health System ("Sistema Único de Saúde" in Portuguese) [2], aligned with the target 3.8 of the Sustainable Development Goals [3]. The National Registry of Healthcare Facilities ("Cadastro Nacional de Estabelecimentos de Saúde"; or CNES, in Portuguese) is a system with the registry of every healthcare facility in Brazil with information on the capacity building and healthcare workforce regarding its public or private nature [4]. The CNES must be fed monthly by each healthcare unit, providing updated knowledge about the existing healthcare network to assist in health planning and effective and efficient management. Public consultations can be made on the CNES system (<http://cnes.datasus.gov.br/>), but the outputs will be provided in separated tables for each system dimension: healthcare facility' general information, physical resources, human resources, services, and health teams. Every dataset dimensions has between 49,989,883 and 55,355,133 registers. Moreover, the number of attributes varies between 451 and 849.

Our objective is to provide a machine-readable and interoperable dataset on the CNES dimensions. This

dataset will support studies that describe the Brazilian healthcare situation, that employ healthcare dimensions as confounders in ecological designs, or studies on the impact of public policies on health outcomes. Data from CNES are currently only available through tabular sets or via FTP and in a disaggregated way. The presented datasets make them available to the community, in a standardized way and following the FAIR principles. They provide unified access to the datasets and enrichments with geospatial information, which allows analysis in total granularity (accessing Brazilian health facilities), and spatial and temporal aggregated data.

Data description

The datasets created contain the different dimensions of the National Registry of Health Facilities (CNES), which are: physical resources (beds and equipments), human resources (health professionals), services and health teams. The dataset and its related files are available for download, as shown in Table 1. Each of these datasets results from a construction method with the following steps: (i) data extraction, (ii) data integration, (iii) data mapping, (iv) reshaping data, and (v) data enrichment. The criteria adopted for data management were based on

detailed studies of the dataset, and the support of specialists in the field.¹

- i. The datasets were extracted from DATASUS, and each one represents a different dimension of health facilities, with records of physical resources, such as types of existing beds and amounts of equipments existing and in use in the health facilities, as well as the health services offered, numbers of professionals and types of health teams. For this, data on health facilities (ST), complementary data from the CNES (DC), and data from each dimension were extracted. In addition, data extraction was performed over time (months and years) and in geographic space (federative units), with the observation of health units.
- ii. The datasets developed are the result of integrating the three types of databases previously extracted: complementary data, health facilities, and dimension, aggregated by competence date (months and years) and federative unit in Brazil.
- iii. The data mapping step aims to create variables where, for each numeric value of a source variable, a categorical definition value is created.
- iv. To reshape data, the values in indexes/columns are organized, creating new datasets. From there, the treatment of invalid values is performed.
- v. Finally, the dataset was enriched with information about the municipalities, states, and regions obtained from the DATASUS FTP server.² The attributes resulting from this process are the coordinates, area, regional health codes, and other Boolean attributes, indicating if the municipality is the state's capital or border, for example. The data dictionary³ presents a full description. All of these steps are explained in details and can be reproduced for the CNES Teams dataset, as an example, with the Jupyter Notebook file available at <https://doi.org/10.7303/syn34388051.1>.

The constructed datasets provide information about the CNES in a machine-readable, ready-to-use format. They allow evaluating and understanding the spatial distribution and temporal evolution of Brazil's private

and/or public health infrastructure. Furthermore, they contribute to the aggregation of ecological studies, with the municipality as the primary unit of analysis, helping in studies that seek to observe particular public health aspects such as hospitalization for diarrhea or pneumonia, vaccine coverage, and preventable infant mortality, for example [17].

Limitations

- As the feeding of this data is the responsibility of the health facilities themselves, errors and inconsistencies in the registration of data may occur, as well as the discontinuity of the same. Thus, despite the data being monthly, its quality is not guaranteed on a monthly basis.
- When using these data, consider that the first years were periods of transition/adaptation of the system, as well as the correct use and registration of health facilities in the CNES, so inconsistencies and duplicities may occur, mainly in the first 3 years.
- There are variables which go into disuse, therefore their feeding are discontinued.
- The year 2005 is included in the dataset from the eighth month onwards, and as it represents the beginning of the CNES system, its data are not fully complete in the first months.

Acknowledgements

The authors thank CNPq, CAPES, and FAPERJ for partially sponsoring this research.

Author contributions

All authors contributed equally to the study. CB, PB and RA conceptualized the study design; VR, RG, VK, BP, GS, CC and LC acquired the data; RS and VR conducted the data Extraction, Transformation, Loading (ETL) process; RS, RG and VR documented the ETL process and wrote the manuscript. Furthermore SC; CL; RF; MM; LZ; GA; AA; NN; JL; MP revised it critically for intellectual content. All authors read and approved the final manuscript.

Funding

This work was supported, in whole or part, by the Bill & Melinda Gates Foundation [Grant ID INV-027961] and National Council for Scientific and Technological Development [Grants 445877/2020-0 and 445882/2020-4]. Under the grant conditions of the Bill & Melinda Gates Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

Availability of data and materials

The CNES datasets corresponding to beds [5], equipments [6], professionals [7], services [8], and teams [9]; the CNES data dictionaries related to beds [10], equipments [11], professionals [12], services [13], and teams [14], as well as the CNES ETL process report [15] and an descriptive example of the CNES teams ETL process [16] is freely and openly available on the Synapse repository at <https://doi.org/10.7303/syn26343275>. Anyone can browse the content on the Synapse website, but you must register for an account using your email address to download the files and datasets. Please see Table 1 for details and links to the data.

¹ Description of the construction of the datasets is available at <https://doi.org/10.7303/syn26343275>.

² Available at <ftp://ftp.datasus.gov.br/dissemin/publicos/SINASC/NOV/TABELAS/>.

³ Available at <https://doi.org/10.7303/syn26343275>.

Declarations

Ethics approval and consent to participate

We used data from open sources. The Brazilian Institute of Geography and Statistics, the Institute for Applied Economic Research, and the Ministry of Health of Brazil are committed to respecting the ethical precepts and ensuring data privacy and security. The Brazilian legislation exempts the use of public and anonymized secondary data from ethical approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 8 December 2022 Accepted: 3 July 2023

Published online: 20 July 2023

References

- World Health Organization. Primary health care: key facts. Technical report (April 2021). 2021. <https://www.who.int/news-room/fact-sheets/detail/primary-health-care>. Accessed Nov 2022.
- Brasil: Lei no 8.080, de 19 de setembro de 1990, Brasília, DF. 1990. http://www.planalto.gov.br/ccivil_03/leis/18080.htm. Accessed Nov 2022.
- United Nations. Sustainable development goals. Technical report. <https://sdgs.un.org/goals>. Accessed Nov 2022.
- Brasil: Portaria No 1.646, de 2 de outubro de 2015, Brasília, DF. 2015. https://bvsms.saude.gov.br/bvs/saudelegis/gm/2015/prt1646_02_10_2015.html. Accessed Nov 2022.
- Dataset CNES Beds. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32209177.1>.
- Dataset CNES Equipments. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32208612.1>.
- Dataset CNES Professionals. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32210090.1>.
- Dataset CNES Services. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32211006.1>.
- Dataset CNES Teams. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32207908.1>.
- Dictionary CNES Beds. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32167629.1>.
- Dictionary CNES Equipments. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32167630.1>.
- Dictionary CNES Professionals. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32167631.1>.
- Dictionary CNES Services. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32167632.1>.
- Dictionary CNES Teams. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn32167633.1>.
- Report of the CNES ETL Process. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn34388050.2>.
- Example of the CNES Teams ETL Process. National Registry of Healthcare Facilities of Brazil (CNES) gathered together by PCDaS. 2022. <https://doi.org/10.7303/syn34388051.1>.
- Fernandes Santos Alves R, de Moraes Mello Boccolini P, Baroni LR, de Almeida Relvas-Brandt L, de Abreu Junqueira Gritz R, Siqueira Boccolini

C. Brazilian spatial, demographic, and socioeconomic data from 1996 to 2020. *BMC Res Notes*. 2022;15(1):1–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

