SEDHI
Unit on Social and Environmental Determinants of Health Inequalities

GOVERNANCE OF INTEGRATED ADMINISTRATIVE DATA FOR RESEARCH IN BRAZIL: THE PIONEERING EXPERIENCE OF CIDACS/FIOCRUZ BAHIA

cidacs
Center for Data and Knowledge Integration for Health

FIOCRUZ | Bahia

CENTER FOR DATA AND KNOWLEDGE INTEGRATION IN HEALTH
(Cidacs/Fiocruz Bahia)

Rua Mundo, nº 121, sala 315 - Parque Tecnológico da Bahia
Edf. Tecnocentro, Trobogy, Salvador/BA

CIDACS COORDINATORS
Pablo Ivan Pereira Ramos
Bethânia de Araújo Almeida

CIDACS SCIENTIFIC COORDINATOR
Mauricio Lima Barreto

TECHNICAL PRODUCTION
Scientific Communication and Dissemination Team

TEXT
Bethânia de Araújo Almeida

LAYOUT/DESIGN
Gilson Rabelo

REVISION
Larissa Costa
Mauricio L. Barreto
Mariana Sebastião
Pablo Ivan P. Ramos

International Cataloguing in Publication (CIP) Data
(Gonçalo Moniz Institute Library/FIOCRUZ – Bahia)

A447p **Almeida,Bethâniade Araújo**

Governance of integrated administrative data for research in Brazil: the pioneering experience of CIDACS/FIOCRUZ Bahia. Layout/Design by Gilson Rabelo; Revision by Larissa Costa; Mauricio Lima Barreto; Mariana Sebastião; Pablo Ivan Pereira Ramos; Technical Production by Scientific Communication and Dissemination Team. _ 1. ed. _ Salvador: CIDACS/Fiocruz Bahia, 2025.

16 f.: il.: 21 cm

ISBN 978-65-01-40242-0

1. Experience report. 2. Administrative data. 3. Data integration. 4. Public health research. I. Rabelo, Gilson. II. Costa, Larissa. III. Barreto, Mauricio Lima. IV. Sebastião, Mariana. V. Ramos, Pablo Ivan Pereira. VI. Scientific Communication and Dissemination Team. VII. Title.

CDU 001.891

CONTENTS

INTRODUCTION

Government administrative data are routinely collected by public agencies to fulfill their core activities. These data contain information on individuals or families who interact with the public sector over time, from conception to death. Characterized by large sample sizes, well-defined populations, and temporality, which permits use in longitudinal studies, these data have great value for the development and evaluation of public policies, as well as for scientific research. Administrative data can be transformed into knowledge, evidence and insight to aid in the understanding of diverse problems, and hold a great capacity to support decision-making and the provision of services to the population.

This type of data is routinely collected and recorded in national information systems to manage administrative activities and provide services to the population. Brazil possesses a diverse array of national information systems to support public administration, with examples including the Ministry of Health's Information Systems, including the Information System on Live Births (Sinasc), the Mortality Information System (Sim), and the Notifiable Diseases Information System (Sinan). In the health sector, mapping conducted between 2010 and 2018 identified 54 different national information systems (Neto e Chioro, 2021), revealing the complexity of the administrative data ecosystem.

These systems are presented and made available in an isolated manner. Data is generally provided in an aggregated and open fashion to promote public transparency, social oversight, and to foment scientific research and innovation. However, when the integration of these data becomes necessary, detailed information — microdata with identification — is required to enable the finding and integration of information on unique individuals across different databases.

The integration of administrative data for population-based studies and research has the potential to generate knowledge to support public policies and improve living conditions in a variety of areas, such as public health surveillance and research (Haneef et al., 2020). For this reason, data integration is a strategy that has been adopted by many

countries, particularly those in the so-called Global North, often through partnerships between governments and research institutions.[1]

In Brazil, the Center for Data and Knowledge Integration in Health (Cidacs/Fiocruz Bahia), was founded in 2016 with the mission of conducting interdisciplinary research focused on social health inequalities using integrated data to produce knowledge capable of supporting the formulation and evaluation of public policies.

The present document details Cidacs' experience with regard to its aims to support evidence-based public policies, and is ultimately designed to inform recommendations for Brazil to advance its production and use of integrated administrative data to conduct research in the public interest.

THE CENTER FOR DATA AND KNOWLEDGE INTEGRATION IN HEALTH (CIDACS/FIOCRUZ BAHIA)

Established as a research laboratory linked to Fiocruz Bahia, Cidacs originated from a project designed to integrate data on health and social assistance, namely the 100 Million Brazilians Cohort, which sought to investigate the impacts of governmental protection policies on social determinants of health in the Brazilian population (Barreto et al., 2021).

All applicants for federal social assistance programs must complete a detailed questionnaire to collect demographic, economic and social information about each family member, in addition to their household characteristics, which is then stored in the Unified Registry (CadÚnico)[2]. This information is used as criteria for the inclusion of families in federal programs, as well as to organize the provision of programs and services targeting Brazilians living in conditions of socioeconomic vulnerability.

CadÚnico is an administrative database that, upon integration with other databases, possesses unprecedented potential due to its extensive coverage of low-income

[1] Some examples in the United Kingdom, Australia and Canada are, respectively, Administrative Data Research UK (https://www.adruk.org/), Australian Institute of Health and Welfare (https://www.aihw.gov.au/our-services/data-linkage) and Population Data BC (https://www.popdata.bc.ca/).
[2] Additional information on CadÚnico is available at: https://www.gov.br/mds/pt-br/acoes-e-programas/cadastro-unico#:~:text=O%20Cadastro%20%C3%9Anico%20para%20Programas,residentes%20em%20todo%20territ%C3%B3rio%20nacional.

populations in Brazil. It contains a wealth of sociodemographic information that underpins investigations into social determinants of health and the effects of social policies and programs on health outcomes over time.

The 100 Million Brazilians Cohort was built by integrating administrative data collected by federal social and health programs. The cohort, which currently consists of over 140 million individuals, uses CadÚnico as its baseline dating back to 2001. Internationally recognized as a differentiated data resource in low- and middle-income countries, the size of the cohort database provides unparalleled statistical potential for hypothesis testing and causal inference analysis (Nissan et al., 2022, p. 42).

The 100 Million Brazilians Cohort has been used to support research on diverse topics, such as the effects of social policies on child and maternal health (Ramos et al., 2021; Alves et al., 2023), social inequalities in health (Rebouças et al., 2022; Bennes et al., 2024), infectious diseases (Pescarini et al., 2020; Paixão et al., 2022; Silva et al., 2024) and mental health (Machado et al., 2022; Toledo et al., 2024), among other topics.

Research results are published in nationally and internationally recognized peer-reviewed journals, demonstrating the suitability of integrated data for methodologies, analyses and conclusions inherent to high-quality science. Cidacs offers visibility and transparency regarding its work through information posted on its website, social media channels, communication materials and activities promoting public scientific engagement.

To enable the creation of the 100 Million Brazilians Cohort, Cidacs' founding members drew on bibliographic research and international experiences, secured national and international funding through research projects and established interdisciplinary collaborations to develop processes, workflows, methodologies and strategies for integrating large volumes of data in a manner suitable for conducting research in the Brazilian context. This was done in accordance with ethical regulations for scientific research involving humans and with robust protections for personal data. These efforts culminated in the construction of the Integrated Data Platform.

THE INTEGRATED DATA PLATFORM

Through an agreement between Fiocruz Bahia and the Bahia State Government, space was secured to set up the Integrated Data Platform in a building, Tecnocentro, belonging to the State Secretariat for Science, Technology and Innovation (Secti).

The platform possesses physical and computational infrastructure developed to receive, process and provide access to high-quality and accurate integrated datasets for population-based research and studies, prioritizing information security and the privacy of data subjects throughout all processes.

The integration of records between different databases was crucial to developing the 100 Million Brazilians Cohort. However, the inclusion of each individual's unique identification number (CPF) as an identifier for Brazilian citizens was only recently implemented under legal regulation (Lei 14.534/2023)[3]. In the absence of a unique identifier, a set of identifying attributes available across different databases is necessary to apply record linkage techniques. All processing throughout the lifecycle of original and integrated data is conducted in a secure and controlled environment, in line with best practices for data integration to address research inquiries (Barreto et al., 2019)[4].

Considering the heterogeneity and large volume of data in question, Cidacs developed a proprietary record linkage (Cidacs – RL) algorithm to identify sets of similar attributes among pairs, enabling the platform to operate among diverse administrative databases with scalability, quality and high accuracy (Barbosa et al., 2020).

Database requests are always based on updates to the 100 Million Brazilians Cohort, considering the needs of each individual research project. Based on each request, a data controller is formally contacted with the appropriate justification for the use of data in a proposed study context, including a copy of ethical approval by an institutional review

[3] Lei 14.534/2023 specifies and establishes the Individual Taxpayer Identification (CPF) number as sufficient to identify citizens in public service databases. Information available at: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14534.htm

[4] Data integration for Cidacs research purposes follows international principles and recommendations for secure research environments, which are only accessible by approved researchers.

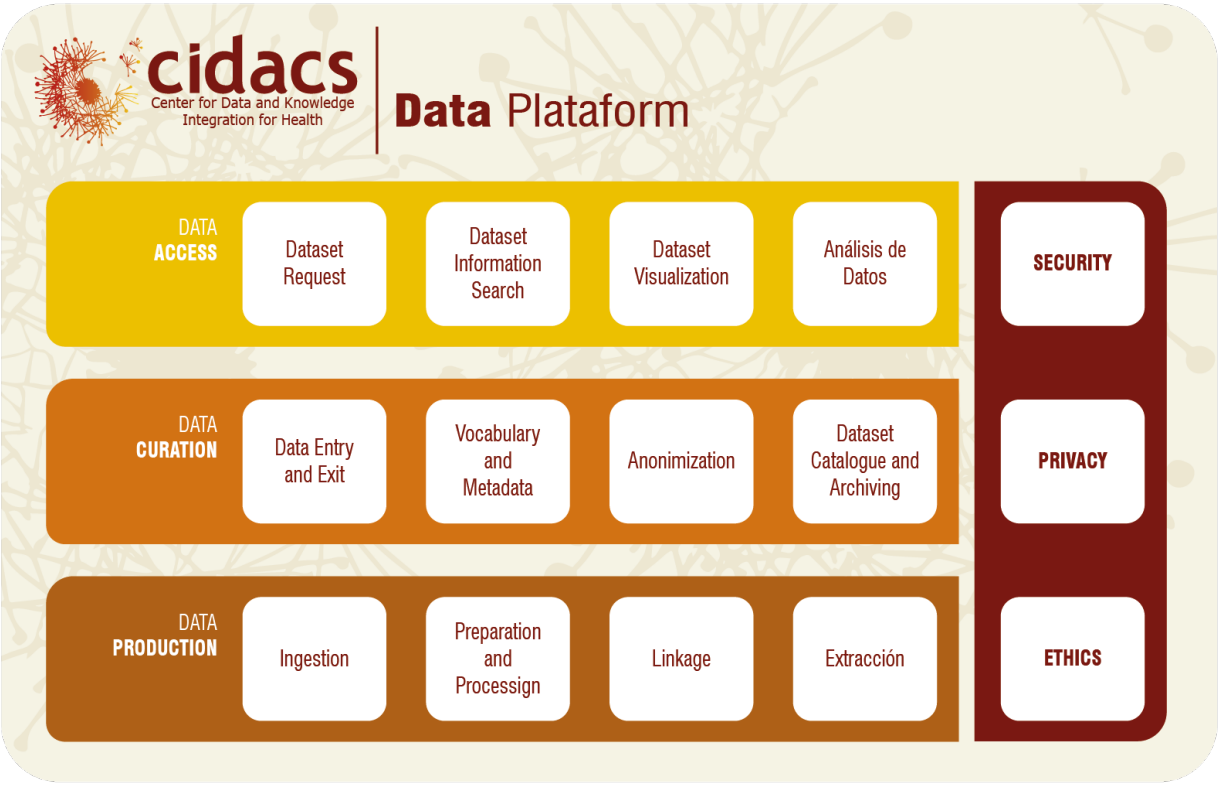
board from the CEP/Conep system[5]. Following the grating of a request, Cidacs follows established workflows that vary from controller to controller[6].

After providing all necessary documentation, copies of datasets are prepared and securely sent to Cidacs. Upon arrival at Cidacs, a term of receipt is prepared, and data is immediately transferred to a secure environment with its own, separate data network, which is disconnected from Cidacs' network and has no internet access. Upon validating consistency, data are cataloged and transferred to a storage location.

The integration of administrative data and transformation into longitudinal structures appropriate for addressing research questions occur in the Secure Room environment, using methodological strategies for selecting, cleaning, standardizing and harmonizing variables. Two scenarios for data linkage are employed: deterministic, involving the linkage of database pairs containing unique identifiers, e.g. NIS or CPF; and non-deterministic, in which the absence of a common unique identifier requires linkage via other individual variables, i.e., name, sex, date of birth, municipality of residence, mother's maiden name, etc.

Pre-processing is an intensive procedure performed on each original database. This involves not only preparing variables for integration but also includes contributions from epidemiologists and statisticians to ensure a useful data format for research purposes. When necessary, the governmental departments that originally provided datasets are consulted to clarify any issues regarding variable information or the format used to code data. From an operational standpoint, pre-processing involves data cleaning, standardization, harmonization and integration, as well as de-duplication and the selection of variables considered useful for research.

[5] The CEP/CONEP System ethically evaluates research protocols involving human beings in Brazil. The system is formed by the National Research Ethics Commission (CONEP) and local Institutional Review Boards (i.e., Research Ethics Committees—CEP in Portuguese).
[6] It is important to recognize gaps in the establishment and dissemination of criteria, requirements and workflows guiding access to government administrative data for academic research and public health studies across the country's government departments.



Source: Cidacs, 2017

Subsequently, after evaluating data quality and accuracy, anonymized or pseudonymized integrated datasets, which contain strictly the variables necessary to address the questions under study, are transferred from the Secure Room to the Access Environment.

Requests for transferring integrated datasets are made under registration and protocol by Cidacs' Curatorship team and executed by the Information Security team. The transfer of integrated data to the access and analysis environment is performed physically on encrypted disks, as the Secure Room is disconnected from other data networks.

Access to and analysis of integrated data to address research inquiries occurs either in-person or exclusively via secure networks (using Virtual Private Network–VPN). After receiving credentials and being properly registered, a research team member will gain access to the Access Environment.

The Access Environment is separate from other computational structures and blocks access to the internet for users connected via VPN. It provides a virtual desktop and pre-installed audited analytical tools for the research team to utilize. Despite being

anonymized, data cannot be downloaded. Only the output of aggregated data (tables, graphs and indicators) is allowed from this environment after submitting a request for results retrieval, which are manually inspected to mitigate the risk of re-identifying data subjects.

IN DEFENSE OF USING INTEGRATED DATA FOR RESEARCH IN BRAZIL

Cidacs was inaugurated in December 2016, prior to the enactment of the Brazilian Data Protection Law (LGPD), which was approved by Brazil's legislature in 2018 and went into effect in 2020. In light of the need to develop processes, workflows and strategies for integrating information across different databases in accordance with ethical regulations for scientific research and personal data protection, it was necessary to analyze both publications in the literature as well as international experiences.

The European Union's General Data Protection Regulation (GDPR) and legal consultancy provided by Danilo Doneda, a personal data protection specialist, were fundamental in establishing the principles and framework that guides Cidacs' use of administrative data for research purposes.

It is noteworthy that in discussions surrounding legislative proposal 4.060/2012, upon which the LGPD was predicated, concerns surrounding personal data processing largely centered on the commercialization and monetization of data, particularly by big tech companies. Considering that personal data can be transformed into commodities or public goods depending on specific interests and usage intents, Professor Mauricio Barreto[7] sought to mobilize the Brazilian scientific community to participate in discussions on data protection.

[7] Mauricio Lima Barreto is the founder of Cidacs/Fiocruz Bahia, professor emeritus of the Federal University of Bahia, senior researcher at Fiocruz Bahia, level 1A researcher at the National Council for Scientific and Technological Development (CNPq), full member of the Brazilian Academy of Sciences and The World Academy of Science (TWAS).

Source: Special commission on legislative proposal 4.060/12 – Processing and Protection of Personal Data | Chamber of Deputies 22/05/18.

On May 22, 2018, Professor Mauricio Barreto participated in a public hearing focused on data protection in the fields of Science and Technology, Communication and Information, promoted by the Special Committee charged with deliberating legislative proposal 4.060/12[8]. On this occasion, he defended the use of personal data for scientific research and studies aimed at producing knowledge to support evidence-based public policymaking, so long as appropriate prevention and security measures are applied to protect data subjects' privacy and rights.

On August 14, 2018, the Brazilian LGPD (Lei n° 13.709/2018[9]) was enacted, thereafter entering into effect on September 18, 2020. It represented a significant milestone in Brazilian legislation, establishing rules for the processing of personal data, whether physically or digitally, by public or private individuals or legal entities.

In its text, the LGPD recognizes academic research and public health studies as specific contexts for processing personal and sensitive data (Articles 4, 7, 11 and 13), provided that the recommendations established in the law are followed.

Article 13 of the LGPD specifically addresses the conduct of public health studies, stipulating that

> In the conduct of studies on public health, research institutions may access personal databases, whose data must be processed exclusively within the institution and strictly for the purpose of conducting studies and research, maintained in a controlled and secure environment, in accordance with specific regulations covering security practices, including anonymization or

[8] Available at: https://www.youtube.com/watch?v=wZHgfHZruHY&t=7346s

[9] Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.html

> pseudonymization whenever possible, and in observation of ethical standards related to studies and research (BRASIL, 2018, Art. 13).

Other paragraphs in this same article establish that the disclosure of results or any part of a study or research endeavor must never reveal personal data; all research institutions will be responsible for information security, never allowing data transfer to third parties; and that access to data will be regulated under national authority as well as by health and sanitation authorities within their specific competencies.

Questions surrounding the interpretation of the LGPD with respect to academic research and public health studies led the National Data Protection Authority (ANPD) to publish a technical study entitled "The LGPD and the processing of personal data for academic and research purposes,"[10] which demonstrated the need for public debate and to receive contributions to inform decisions regarding the access to personal data and processing for academic and research purposes with legal certainty and respect for the rights of data subjects.

The ANPD study cites an article, published in 2019 by Mauricio Barreto, Bethânia Almeida and Danilo Doneda (ANPD, 2022, p. 10), stating that research activities, considered a specific context for processing personal data, must balance individual rights versus pursuits in the public interest through the application of sufficient and suitable technical and organizational measures to ensure data protection as well as the rights of data subjects[11].

Contributions to the technical study, subjected to public consultation and analysis by ANPD, resulted in the publication of a document entitled "Guidelines for Processing Personal Data for Academic and Research Purposes"[12].

The nature of the work undertaken by Cidacs and the entry into force of the LGPD gave rise to a research project entitled "Perceptions and experiences on data sharing and linkage for research and the evaluation of public health policy", led by a CIDACS researcher (Dr. Bethânia Almeida). This exploratory study focusing on a novel topic in Brazil aimed to understand the experiences and perceptions of individuals across

[10] Records of public deliberation and access to the ANPD's technical study are available at: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-estudo-tecnico-a-lgpd-e-o-tratamento-de-dados-pessoais-para-fins-academicos-e-para-a-realizacao-de-estudos-por-orgao-de-pesquisa

[11] BARRETO, M.; ALMEIDA, B.; DONEDA, D. Uso e proteção de dados pessoais na pesquisa científica. RDU, Porto Alegre, Volume 16, n. 90, nov-dez 2019, p. 189. Available at: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3895/Doneda%3B%20Barreto%3B%20Almeida%2C%202019

[12] The ANPD publication (2023) guiding the processing of personal data for academic purposes and the conduct of studies and research is available at: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/materiais-educativos-e-publicacoes/web-guia-anpd-tratamento-de-dados-para-fins-academicos.pdf

different segments of society with regard to the sharing and linkage of personal data, in both public and private spheres, with special attention paid to government administrative data (Almeida; Pimenta, 2021).

In an effort to follow applicable ethical standards for conducting studies and research, Cidacs/Fiocruz Bahia submitted contributions to the National Research Ethics Committee (Conep) during public consultations on the use of databases in research involving humans, which took place in September 2023[13]. Following the publication of Resolution no. 738 by the National Health Council, which details guidelines on the use of databases for scientific research involving humans, a public hearing with Conep was requested to clarify issues surrounding the use of secondary data containing personal information for scientific research purposes.

By facing challenges in the development of strategies aimed at transforming administrative data into data suitable for research purposes, Cidacs has contributed to national debates on the use of personal data for scientific research and public health studies.

A systematic review and thematic analysis of articles published between 2012 and 2021 on relevant characteristics enabling the transformation of administrative data into data suitable for research purposes identified 38 publications from countries including Australia, Brazil, Canada, China, New Zealand, Taiwan, the United States and the United Kingdom. The search returned just one Brazilian publication on the creation and establishment of Cidacs, detailing its efforts to integrate administrative data for research purposes (Mc Grath-Lone et al., 2022, p. 5-6).

Cidacs' contributions to public health and the integration of data for population-based studies have received international recognition. In 2023, the Center received the Information and Data Distribution Award from the World Federation of Public Health Associations (WFPHA)[14] for its work on the Rede CoVida network[15]. In 2024, Professor Mauricio Barreto, currently Cidacs' Scientific Coordinator, was elected to serve as a member of the executive committee of the International Population Data Linkage Network from 2025-2026[16].

[13] Available at: https://www.gov.br/conselho-nacional-de-saude/pt-br/assuntos/noticias/2023/setembro/conep-consulta-publica-sobre-uso-de-bancos-de-dados-em-pesquisas-com-seres-humanos

[14] https://www.wfpha.org/wfpha-awards-2023/
[15] https://cidacs.bahia.fiocruz.br/plataforma/rede-covida/
[16] https://ipdln.org/introducing-the-2025-2026-ipdln-executive-committee/

INTEGRATED ADMINISTRATIVE DATA GOVERNANCE FOR RESEARCH

The governance of integrated administrative data involves the establishment of policies, processes, standards and technologies to ensure the integrity, quality, privacy, security and availability of this data.

Data governance is complex and highly dependent on specific contexts and different types of data, which require specific methods of organization. Governance requires planning and the alignment of resources and competencies for its implementation in order to achieve objectives in compliance with established ethical, legal, and operational regulations, while aiding in the identification, evaluation and mitigation of risks.

At Cidacs, collaborations and investments in computational resources, personnel and processes were all crucial to creation of the Integrated Data Platform, which enables the receipt, processing, integration, preservation of and access to data for high-quality research conducted in compliance with ethical and legal regulations, as well as the provision of secure data integration environments for research needs.

Cidacs' data management and governance structure was designed to preserve data privacy and confidentiality through physical and cybersecurity measures. Additionally, data management and governance involve collaboration with epidemiologists, data scientists, statisticians, social scientists among others to shed light on population coverage and database limitations, aiding in the development of appropriate methodological strategies to ensure research data quality and utility.

Funding for research activities and other resources for maintaining the Integrated Data Platform and the administrative and operational processes of the Center have been secured through national and international agencies interested in supporting investigations aimed at evaluating the impact of social policies on health, social determinants of health, and, more recently, the effects of climate and the environment on health. Resources are obtained through the submission of research projects to funding agencies by researchers affiliated with the Center, a situation that directly impacts the activities of technical teams and the maintenance of computational infrastructure due to the variable duration of research projects and uncertainties regarding the acquisition of new funding, especially given the unstable funding landscape for research in both national and international spheres.

The governance of integrated administrative data at Cidacs involves the articulation of scientific, technical, ethical, legal and social issues (Almeida et al., 2024) via a dynamic process to advance our understanding of how diverse aspects interact and influence the health of specific groups or entire populations.

Although Cidacs' experience is specific, the challenges the Center has faced allow for recommendations to be made for Brazil to advance its use of integrated administrative data for population-based research to produce results capable of supporting evidence-based public policies.

CONCLUSIONS AND RECCOMENDATIONS

- Recognize the importance and great potential of integrated government administrative data in generating scientific knowledge and evidence-based public policymaking.
- Promote training opportunities and awareness campaigns highlighting the importance of consistency and completeness in the information contained in administrative records, as data quality is impacted by the social nature of data collection.
- Regulate and divulge requirements and workflows granting access to administrative data for research purposes across municipal, state and federal government departments/agencies.
- Establish collaborations between government and academia to articulate knowledge and resources in an effort to promote a national ecosystem of integrated administrative data for research.
- Strengthen Brazil’s capacity for integrating administrative data for population studies in the public interest through the creation of a national center with dedicated human and financial resources and robust governance, while adhering to legal and regulatory frameworks that ensure the rights of data subjects.

REFERENCES

ALMEIDA, B. de A., CARREIRO, R., SOUZA, M. and BARRETO, M. (2024) “CIDACS’ efforts towards an inclusive and dialogic data governance in Brazil: a focused literature review”, *International Journal of Population Data Science*, 9(1), 2024. Available at: https://doi.org/10.23889/ijpds.v9i1.2163.

ALMEIDA,B; PIMENTA,D. Percepções e experiências sobre compartilhamento e vinculação de dados para pesquisa e avaliação de políticas públicas na área da saúde. Salvador, Cidacs, Fiocruz, 2021. Available at: https://cidacs.bahia.fiocruz.br/wp-content/uploads/2021/11/Relatorio7-21.pdf

ALVES, F. J. O. et al. Association of conditional cash transfers with maternal mortality using the 100 Million Brazilian Cohort. *JAMA Network Open*, v. 6, n. 2, p. e230070, 2023. Available at: https://doi.org/10.1001/jamanetworkopen.2023.0070.

BARBOSA, G. C. G. et al. CIDACS-RL: a novel indexing search and scoring-based record linkage system for huge datasets with high accuracy and scalability. *BMC Medical Informatics and Decision Making*, v. 20, n. 1, 9 Nov. 2020. Available at: https://doi.org/10.1186/s12911-020-01285-w

BARRETO, M. L. et al. The Center for Data and Knowledge Integration for Health (CIDACS). *International Journal of Population Data Science*, v. 4, n. 2, 20 Nov. 2019. Available at: https://doi.org/10.23889/ijpds.v4i2.1140

BARRETO, M. L. et al. Cohort Profile: The 100 Million Brazilian Cohort. *International Journal of Epidemiology*, v. 51, n. 2, p. e27–e38, 14 Sep. 2021. Available at: https://doi.org/10.1093/ije/dyab213

BENNES, H. et al. Ethnoracial disparities in childhood growth trajectories in Brazil: a longitudinal nationwide study of four million children. *BMC Pediatrics*, v. 24, n. 1, 10 Feb. 2024. Available at: https://doi.org/10.1186/s12887-024-04550-3

BRASIL. Autoridade Nacional de Proteção de Dados. Ministério da Justiça e Segurança Pública (org.). Guia Orientativo: A LGPD e o tratamento de dados pessoais para fins acadêmicos e para a realização de estudos por órgão de pesquisa. 2022. Available at: https://www.gov.br/anpd/pt-br/assuntos/noticias/sei_00261-000810_2022_17.pdf.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados. Diário Oficial da União, Brasília, DF, 14 Aug. 2018. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

HANEEF, R. et al. Innovative use of data sources: a cross-sectional study of data linkage and artificial intelligence practices across European countries. *Archives of Public Health*, v. 78, n. 1, 10 Jun. 2020. Available at: https://doi.org/10.1186/s13690-020-00436-9

MACHADO D.B., et al. The impact of social drivers, conditional cash transfers and their mechanisms on the mental health of the young; an integrated retrospective and forecasting approach using the 100 million Brazilian Cohort: A study protocol. *PLoS One*. 16 Oct. 2022; 17(10):e0272481. Available at: https://doi.org/10.1371/journal.pone.0272481

MC GRATH-LONE L.M., JAY M.A., BLACKBURN R., GORDON E., ZYLLBERSZTEJN A., WILJAARS L., GILBERT R. What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. *International Journal of Population Data Science*. 27 Apr 2022;7(1):1718. Available at: https://doi.org/10.23889/ijpds.v7i1.1718

NETO, G.C.C; CHIORO, A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? Cadernos de Saúde Pública, 2021; 37(7), e001811. Available at: https://www.scielo.br/j/csp/a/RzNmvjHqmLhPHZp6gfcdC6H/abstract/?lang=pt

NISSAN, H.; DIGGLE, P.; FRONTERRE, C. Combining climate and health data: challenges and opportunities for longitudinal population studies. Wellcome Trust, 2022. Available at: https://cms.wellcome.org/sites/default/files/2022-01/climate-lps-landscaping-report.pdf?_gl=1.

PAIXÃO E.S., et al. Mortality from Congenital Zika Syndrome - Nationwide Cohort Study in Brazil. *The New England Journal of Medicine*. 24 Feb. 2022; 386(8):757-767. Available at: https://doi.org/10.1056/NEJMoa2101195

PESCARINI, J. M. et al. Effect of a conditional cash transfer programme on leprosy treatment adherence and cure in patients from the nationwide 100 Million Brazilian Cohort: a quasi-experimental study. *The Lancet Infectious Diseases*, v. 20, n. 5, p. 618–627, May 2020. Available at: https://doi.org/10.1016/ S1473-3099(19)30624-3

RAMOS, D. et al. Conditional cash transfer program and child mortality: A cross-sectional analysis nested within the 100 Million Brazilian Cohort. *PLOS Medicine*, v. 18, n. 9, p. e1003509, 28 Sept. 2021. Available at: https://doi.org/10.1371/journal.pmed.1003509

REBOUÇAS, P. et al. Ethnoracial inequalities and child mortality in Brazil: a nationwide longitudinal study of 19 million newborn babies. *The Lancet Global Health*, v. 10, n. 10, p. e1453–e1462, 1 Oct. 2022. Available at: https://doi.org/10.1016/S2214-109X(22)00333-3

SILVA, A. F. et al. Income determines the impact of cash transfers on HIV/AIDS: cohort study of 22.7 million Brazilians. *Nature Communications*, v. 15, n. 1, p. 1307, 12 Feb. 2024. Available at: https://doi.org/10.1038/s41467-024-44975-z

TOLEDO L., et al. Risk of psychiatric hospitalization in low-income youth: longitudinal findings from the 100 Million Brazilian Cohort. *International Journal of Epidemiology*, v. 53, n. 6, Dec. 2024, dyae153. Available at: https://doi.org/10.1093/ije/dyae153


cidacs
Centro de Integração de Dados
e Conhecimentos para Saúde
FIOCRUZ | Bahia
cidacs_fiocruz
fiocruzbahia.cidacs
cidacs.fiocruz
CidacsFiocruz
Cidacs-Fiocruz
www.cidacs.bahia.fiocruz.br | Parque Tecnológico da Bahia | Edf. Tecnocentro, R. Mundo, 121 - Trobogy, Salvador - BA, CEP.: 41745-715.