

## CIDACS' efforts towards an inclusive and dialogic data governance in Brazil: a focused literature review

Bethânia de A. Almeida<sup>1,\*</sup>, Roberto P. Carreiro<sup>1</sup>, Maíra L. de Souza<sup>1</sup>, and Mauricio L. Barreto<sup>1</sup>

### Submission History

Submitted:	29/04/2023
Accepted:	20/02/2024
Published:	12/03/2024

<sup>1</sup>Centre for Data and Knowledge Integration for Health (CIDACS), Gonçalo Moniz Institute, Oswaldo Cruz Foundation (FIOCRUZ), Salvador, Bahia

### Abstract

The Centre for Data and Knowledge Integration for Health's (CIDACS) data governance efforts have primarily focused on legal, technical and operational procedures to provide high-quality linked administrative data for investigations on social determinants of health and the impact of social protection policies in low-income and vulnerable populations throughout Brazil. The Centre is moving towards an updated data governance model that incorporates the participation of, and consultation and dialogue with, data stakeholders, including groups covered by our linked data. To this end, this paper presents our procedures and challenges, outlining relevant considerations based on a focused literature review that aims to support the inclusion of societal participation in our revised data governance approach, which should be considered an ongoing process.

### Keywords

data governance; administrative data linkage; data subjects rights as groups and communities

\*Corresponding Author:

Email Address: [bethania.almeida@fiocruz.br](mailto:bethania.almeida@fiocruz.br) (Bethânia de A. Almeida)



## Introduction

The Centre for Data and Knowledge Integration for Health (CIDACS) was created in December 2016 in the city of Salvador (Bahia-Brazil) to conduct interdisciplinary research on population health using integrated Brazilian (national-level) databases to generate scientific knowledge and provide evidence to support public policymaking. The linkage of social and health administrative data is at the core of the centre's activities [1]. CIDACS has recently started incorporating environmental and climate data to link with existing health and social administrative data. To date, very few Brazilian studies have employed data linked on an individual level.

The usage of secondary data containing personal information for research purposes is restricted in many countries. In Brazil, the legal framework for regulating the use of personal data and preventing abuse is the Brazilian General Data Protection Law (termed Lei Geral de Proteção de Dados Pessoais – LGPD in Portuguese), which has been in effect since 2021 after being passed in 2018. The law applies to any processing of personal data carried out by an individual or public or private legal entity throughout the national territory.

According to the LGPD, the consent of data subjects or their legal guardians is required for the collection, processing and use of personal data, together with guarantees of transparency, security and the minimized use of this type of data. Guaranteeing transparency refers to providing information to data subjects regarding the processing of their data.

In compliance with the law, those responsible for the processing of personal data are required to organise and maintain records about all activities related to the processing of any data for which they are responsible. In addition, the National Data Protection Authority can request information and report on personal data processing to provide oversight in adherence with the LGPD [2].

Considering CIDACS' use of secondary data, it is important to note that the LGPD cites academic research and public health studies as legitimate applications for processing secondary data containing personal information concerning public interests, as long as appropriate ethical, legal, and security measures are implemented. In this context, the legal basis for access, appropriate security arrangements, exclusive usage for a previously specified purpose, appropriate credentials from the requesting institution and the ethical basis of all proposed research endeavours must be evaluated.

## CIDACS data governance: a continuous and cumulative process

### Current procedures

Following the identification of data required to support existing or new studies, when the need arises to use data on an individual level for record linkage, each access request is considered on a case-by-case basis under the workflow determined by data controllers. The acquisition of administrative databases by CIDACS involves a series of negotiations with relevant governmental departments/agencies.

Upon authorisation and under specified conditions, the requested administrative data generated by government agencies are received in a secure, controlled environment not connected to the Internet, in which only authorized personnel are allowed to verify data integrity and initiate data management activities.

Our data governance practices conform to safe data linkage principles and guidelines, including separating linkage and analysis processes [3]. Data is exclusively accessed in a secure environment, also known as a data safe haven or trusted research environment, in adherence with the Five Safes framework (safe people, projects, settings, data and outputs) [4, 5].

CIDACS' data management structure has been designed to preserve confidentiality through physical and cybersecurity measures to guarantee the privacy of identifiable data, e.g. preventing unauthorised access and mitigating the risks associated with data breaches and misuse.

Researchers are only allowed to access pseudonymised linked data containing relevant variables to achieve their proposed study objectives. They must provide written acknowledgement of the terms of responsibility regarding accessing and using CIDACS' data.

Persons who wish to receive authorisation must:

- i) Be affiliated with our institution or be identified as a collaborator;
- ii) Present detailed research projects together with ethical approval by an appropriate institutional review board;
- iii) Provide a CIDACS-approved data plan to guide the linkage and extraction of relevant variables available in the datasets, while restricting these to only what is necessary to satisfy the objectives of the study proposal.

All pseudonymisation of CIDACS' datasets is conducted in a secure environment. This procedure involves the substitution of direct and indirect identifiers with hash codes, ensuring that each dataset produced is unique and non-linkable. Data suppression and generalisation techniques are also applied to reduce data granularity, while providing the variables necessary for each study's analysis, e.g. the corresponding year is the only information available regarding an individual's date of birth.

Researchers perform data analysis via CIDACS' data analysis environment, a safe and secure infrastructure that provides virtual machines with analysis tools that can be accessed in person or remotely. Results obtained from the analysis are available upon request in table, graphic or script format and are only provided following a risk assessment, ensuring that data subjects cannot be re-identified.

## Updating CIDACS' data management procedures

A core factor related to using and reusing linked datasets is aligning our data management procedures with the FAIR principles. This acronym refers to findable, accessible, interoperable and reusable data management practices [6, 7]. To this end, CIDACS aims to provide metadata with unique

and persistent identifiers that are easily discoverable regardless of access rights.<sup>1</sup>

Efforts towards data interoperability are also taking place at CIDACS, including establishing a common data model aimed at addressing the significant challenges associated with standardising variations in data terminologies and formats, as well as in semantics, across a variety of databases [8].

In addition to the continuous process in which data are properly acquired, ingested, prepared and processed to provide linked datasets for research, which is time- and resource-consuming in terms of infrastructure and capacity building, CIDACS has also been developing and consolidating its data management procedures as well as its data governance framework.

## Updating CIDACS' data governance framework

Our current data governance policies and protocols are modelled on organisational structures and procedures based on technical, ethical and legal aspects to support data acquisition, data management and use for public health research purposes. We believe that the CIDACS data governance model has been functioning well in support of our purpose, as it is in adherence with key elements of safe and quality data linkage, as well as access for scientific research purposes.

On the other hand, we also believe that a data governance approach must remain open to addressing key questions that ensure that it is in line with its institutional values and mission. In our specific case, the mission entails the generation of scientific knowledge and provision of evidence to support public policymaking aimed at tackling health inequalities in Brazil.

In addition to the responsibility of promoting and guaranteeing the security of data, with safe linkage and usage for public health research purposes aligned with legal, ethical and privacy requirements, CIDACS recognises the importance of adopting an updated data governance model that incorporates societal components, such as social values and stakeholders interests to enhance public trust and foster cooperation.

## CIDACS and the 100 million Brazilians cohort

The CIDACS data centre was initially constructed to house the 100 Million Brazilians Cohort (N = 131,697,800 as of 2018), conceived to investigate the impact of social protection policies on health in low-income and vulnerable populations throughout Brazil [9].

The cohort consists of all individuals who have applied for any governmental social welfare assistance since 2001 and are thusly registered in a federal government database called the Unified Registry for Social Programmes (CadUnico). Eligibility for registration in CadUnico is contingent upon an

income of up to half the Brazilian minimum monthly wage (approximately USD 125 in 2022) or a total family income not exceeding the equivalent of three minimum monthly wages (approximately USD 750 in 2022).

Applicants answer a detailed questionnaire to collect demographic, economic, and social information on each family member and household characteristics; to remain eligible, each individual's records must be updated every two years. By 2018, 61% of the entire Brazilian population was registered in CadUnico, enabling representativeness deemed suitable for linkage with other administrative databases<sup>2</sup>. Administrative data differs from populational sampling, as these data represent specific population segments that share one or more common attributes. In the case of CadUnico, common attributes are related to an individual's or their family's total income.

Epidemiological studies involving the 100 Million Brazilians Cohort have been comparing disease response and child mortality, among other factors, between groups of recipients and non-recipients of specific social protection programmes, e.g. the Brazilian conditional cash transfer program *Bolsa Familia* [10–12]. CIDACS' work is based on investigating social determinants of health, which must consider the specific conditions in which people are born, raised and grow old [13]. The non-medical factors that influence conditions of daily life and health outcomes require action by all sectors with recognition of health and equity as core responsibilities of governments to its people [14].

In this context, it must be acknowledged that Brazil is not only a country of continental dimensions, but also one of the most inequitable in the world. It is of utmost importance to understand that the CadUnico database contains data on the most impoverished individuals and families living in Brazil, including traditional groups, e.g. groups of different racial/ethnic origins, as well as groups facing extreme vulnerability, such as people experiencing homelessness.

## Administrative data linkage: sociotechnical challenges

In general, the administrative data collected for operational purposes requires a great deal of effort to be "transformed" into data suitable for scientific research purposes, e.g. the mitigation of inaccuracies, accounting for missing data and data harmonisation pose significant challenges to data integration [15].

Furthermore, data standardisation and poor data quality can exacerbate information gaps, including the under representation of specific groups, which renders their hardship invisible. For instance, absent or poor quality data on race and ethnicity collected by national health surveys and vital statistics registries hampers the documentation and monitoring of inequities among Afro-descendent and indigenous populations throughout many Latin American countries, hindering the development and implementation of proposals and policies to address these inequities [16].

Considering that CadUnico is a unique administrative database in Brazil with high population coverage, including

<sup>1</sup>CIDACS is in the process of providing metadata with persistent identifiers from its extracted datasets, available at <https://dataverse.cidacs.org/>.

<sup>2</sup>Cidacs has recently received CadUnico databases corresponding to years 2019, 2020 and 2021, which are undergoing preprocessing to be linked with the existing 100 Million Brazilians Cohort data (2001–2018).

high coverage of specific social and traditional subgroups, there exists great potential for linkage between this database and other sources of data to aid in the investigation of social determinants of health within subgroups [17]. Therefore, to gain insights into, and answer research questions related to, causal or protective factors involved in health outcomes relevant to a given population under study, particularly when information is limited or lacking, it must also examine and acknowledge sociohistorical and cultural factors that shape inequities faced by those groups, e.g. indigenous health in Brazil [18].

To support our data linkage process, CIDACS has developed two tools for different types and sizes of databases [19, 20]. Currently, we are attempting to link more individual-level data on vulnerable populations to the existing 100 Million Brazilians Cohort, which poses additional methodological challenges for linkage error quantification, especially in under-represented groups.

It is important to note that research analysis is focused on a population perspective, while ethical and legal standards, as well as privacy preservation, target individuals. CIDACS operates in the realm of population data science, which can be succinctly defined as the science of data about people, characterised by the use of data in a positive manner to benefit citizens and society, as well as efforts supported by data integration from multiple sources. The obtained results are then analysed from a populational perspective, requiring technical infrastructure in conformity with local ethical and legal standards for scientific research purposes while preserving privacy [21].

Due to administrative data linkage challenges, we believe that our novel data governance approach should address issues related to data ethics as well as the rights and interests of the data groups covered by our linked data. We believe that developing this new framework can offer opportunities to enhance and retain public trust in our data practices, particularly with respect to vulnerable and marginalised populations.

Taking into account CIDACS' aspirations, we have investigated publications and initiatives considered relevant to support our efforts to achieve these objectives. The purpose of this paper is to present CIDACS' efforts as a case study, detailing a data centre in a Latin American country currently undergoing the process of updating its data governance approach.

## Methods

We pursued an exploratory approach to identify and analyse multiple initiatives and publications considered important to provide insights to enhance the process of updating CIDACS' data governance approach. Considering that the Centre has implemented mechanisms aligned with key elements of good data governance practices regarding safe data linkage for scientific research, this article does not attempt to offer a systematic or scoping review of data governance elements or framework design. Our investigation also was not systematically restricted with respect to publication date or a specific literature review strategy. Instead, it was rather a focused literature review

as publications were selected based on relevant concepts and characteristics adherent to the inclusion of societal components in the governance of linked administrative data. While the findings and discussion presented here are highly specific to our case, some aspects are more generalisable to other contexts. The discussion section presents CIDACS' initial efforts related to our results, and finally proposes next steps towards implementing an updated data governance model.

## Results

### Data subjects and societal considerations

Administrative data linkage for scientific research involves combining data from various administrative sources, such as health records, census data, and educational records to gain insights and perform research. While this can be immensely valuable for generating scientific knowledge and informing policymaking decisions [22], it also raises several important societal issues, which require robust data governance to ensure the responsible and ethical usage of linked data.

The public's support and acceptance of linked administrative data for scientific research depends on transparency, confidentiality, data security, trust and independent oversight. Public concerns include administrative data access by the private sector, accessing data to obtain profit (i.e. the commercial use of data) and fears surrounding government usage of citizens' data that could result in harm [23, 24].

To ensure the public's support with respect to linking administrative data for scientific research purposes, it is essential that both individuals and society as a whole recognise the personal and common benefits that can be derived from the usage of their data [25]. This becomes even more evident in the context of health research and the elaboration of public policies aimed at improving national health systems and maximising the societal benefits of research, while mitigating potential harm to individuals or groups whose data is being linked [26].

### Technical, ethical and legal considerations

Developing and applying linkage methodologies are essential to enhance quality and avoid potential bias. The methods used for data linkage should be transparent and well-documented to provide reliable linked data for research that may serve to inform evidence-based policymaking [27].

Ethical considerations on administrative data linkage for scientific research are much more related to privacy protection than data subjects' autonomy, as using this type of secondary data implies considerable challenges regarding the retroactive obtainment of consent. In this sense, paramount measures have been taken to implement robust data security [28, 29], as well as the application of techniques related to anonymisation to mitigate the risk of an individual's re-identification [30, 31].

These aspects are in line with the ethical principles followed by academics, as well as laws such as the Brazilian LGPD and EU General Data Protection Regulation (GDPR). These laws were designed to provide an appropriate framework to protect

fundamental individual rights, privacy and confidentiality in accordance with ethical research standards [2, 32]. It is important to note that ethical and regulatory guidelines are generally centred on the security and confidentiality of data subjects as individuals, as specified in legislation aimed at personal data protection.

Considering that CIDACS operates in the realm of population data science, involving the analysis of integrated data from a populational perspective [21], the rights of data groups must be prioritised in the process of updating our data governance structure.

## Data groups' rights and the governance of linked administrative data

Detailed data profiles could potentially lead to stigmatisation or discrimination against certain groups based on characteristics revealed through linkage, particularly when working with sensitive and vulnerable populations [33]. Some findings in the literature and reports of experiences on data governance itself highlight the need for ethical and regulatory frameworks to extend the rights of individual data subjects, groups and communities beyond technical, physical and organisational safeguards and controls designed to protect personal and sensitive data processing, linkage, access and preservation [34–36].

It follows that ethical considerations related to data should be expanded. Floridi and Taddeo (2016) highlighted some ethical problems posed by the collection and analysis of large datasets and the use of big data in research, as well as the ethics of algorithms, the practices of people and organizations in charge of data processes, and the strategies and policies implemented to protect the rights of individuals and groups [37].

Considering data groups' rights, indigenous data sovereignty movements, including the CARE Principles for Indigenous Data Governance and the SEEDS Principles, highlight important fundamental aspects, such as the rights of data subjects as individuals, groups, communities, populations, and even countries over their data. These movements and initiatives suggest a data governance approach that not only recognises cultures and identities in data representativity but further considers data governance as a contextualised and continuous process to be established by each indigenous nation, supported by deliberation to mitigate asymmetries of power and promote social justice [38–41].

Responsible and inclusive data governance approaches must endeavour to mitigate asymmetries and promote equity, necessitating a shift away from communication and consultation towards involving people in the use of data to enable people and society to influence and shape the data governance process [42].

Some initiatives that use administrative data and data linkage to support scientific research have designed strategies to promote dialogue and involvement with society by establishing public panels, public events and community representative panels [43]. The SAIL databank pioneered the establishment of a consumer panel in 2011 to allow the public to voice their views on the databank's work and associated initiatives [44]. In addition, SAIL created a policy detailing

guiding principles on public involvement and engagement to support safe, ethical and responsible data governance in the field of Population Data Science [45].

From a social science perspective, data governance models must consider aspects related to stakeholder interests and reciprocity, governance goals, value from the data and governance mechanisms. Moreover, these stakeholder roles, interrelationships, the articulation of values and the organisation of governance principles, instruments and functioning, are all intertwined and may potentially impact how data is accessed, controlled, used and benefited from [46].

## Discussion

### CIDACS' initial efforts to update its data governance structure

Our literature review indicated the importance of better understanding the public's attitudes towards the usage of linked administrative data. The public's support for, and acceptance of, linked administrative data usage are highly dependent on transparency, confidentiality, data security, the reputation of the organisations performing linkage, and a recognition of personal or common benefits that can be derived from the usage of citizens' data. Fears surrounding the potential harmful use of data by governments, particularly that pertaining to the poorest and most vulnerable individuals and groups, must be taken into account. The results of an investigation previously conducted by CIDACS are convergent with the findings in the literature investigated herein [47].

It is important to highlight that our findings in the literature call attention to the fact that some characteristics revealed through data linkage, particularly when working with sensitive and vulnerable populations, must be considered to mitigate risks of discrimination or stigmatisation against certain groups. A core factor relevant to our situation is that CIDACS acts as a custodian of a database containing data on the lowest-income individuals and families in Brazil; the centre is also attempting to link information on additional subgroups to the existing 100 Million Brazilians Cohort. Accordingly, it is important to consider specific groups' rights, interests and concerns, as well as their cultural specificities and demographic differences in the data governance decision-making that informs scientific research and public policy.

Another relevant finding of our review was the development and application of linkage methodologies aimed at enhancing quality and avoiding potential bias. This has prompted us to start working on developing new methods for bias measurement to evaluate the impact of data linkage quality on the results of observational studies.

Additionally, the literature indicates that ethical issues on data linkage are highly related to robust data information security to safeguard individuals' confidentiality, which aligns with our current data governance approach. Moreover, some findings in the literature also raise concerns that ethical considerations should be expanded to support the design and implementation of responsible, inclusive and contextualised data procedures, policies and practices, which supports our objective of updating the CIDACS data governance framework.

We identified some initiatives using administrative data and data linkage to support scientific research that designed and implemented strategies to promote dialogue, involvement and engagement with society in an effort to give the public an opportunity to voice their views. Likewise, CIDACS has undertaken some initiatives to make the public aware of the terms, conditions and purposes of using data containing personal information and linked administrative data for scientific research. These aim to present CIDACS' work to a diverse range of audiences and to encourage public involvement and participation in research [48, 49].

Currently, we are conducting studies on ethical, legal and societal issues related to data linkage for scientific research and public health research purposes in Brazil and Latin America to better understand both our country's and the overall region's landscape and specificities, since the literature mostly contains experiences, written in English, on data governance pertaining to linked data in high-income countries.

## Next steps towards implementing our updated data governance

In an attempt to update our data governance to be more dialogic and inclusive in the Brazilian context, we are in the process of delineating guiding principles and currently expect to present our proposed revisions to a participatory data governance panel in 2024/2025. This panel will consist of data stakeholders (e.g. data controllers, curators, scientists, users and data subjects), notably including the members/representatives of data groups covered by our linked data to encompass a broad range of views. Their participation should enhance transparency and promote inclusion, as well as foster public trust and cooperation with our data practices.

Establishing a participatory data governance panel is considered a crucial step in the updating process. Beyond presenting and discussing our proposals to support effective decision-making with regard to our updated model, we considered it a way to recognise and acknowledge the data stakeholder contributions that make CIDACS' work with administrative data linkage possible.

## Funding statement

CIDACS has received support from the Wellcome Trust, the Global Health Research Unit of the National Institute for Health and Care Research (NIHR) on Social and Environmental Determinants of Health Inequalities (SEDHI), Bill and Melinda Gates Foundation and Health Surveillance Secretariat, Ministry of Health, Brazil.

## Acknowledgements

We thank the anonymous reviewers and the editors for their constructive feedback that significantly improved the article.

## Conflicts of interest

Despite being employed by the Centre for Data and Knowledge Integration for Health, the authors declare no existing conflicts of interest with regard to using CIDACS as a case study.

## Ethics statement

This study did not require ethical approval as no data subjects were directly involved.

## Data availability statement

No data has been made available. All materials used to support the information in the manuscript are provided in footnotes and references.

## References

1. Barreto ML, Ichihara MY, Almeida BA, *et al.* The Centre for Data and Knowledge Integration for Health (CIDACS): Linking Health and Social Data in Brazil. *International Journal of Population Data Science*. November 2019. <https://doi.org/10.23889/ijpds.v4i2.1140>
2. Brazil. Law on Treatment and Protection of Personal Data: Law 13.709, From 14 August 2018, Modified by Law 13.853, from 8 July 2019. Retrieved from: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm).
3. Harron K, Dibben C, Boyd J, *et al.* Challenges in administrative data linkage for research. *Big Data & Society*. 2017 4(2). <https://doi.org/10.1177/2053951717745678>
4. UK Health Research Alliance & NHSX. Building Trusted Research Environments: Principles and Best Practices; Towards TRE ecosystems (1.0). Zenodo, 2021. <https://doi.org/10.5281/zenodo.4594704>
5. AIHW – Australian Institute of Health and Welfare. Data Governance Framework, 2021. Retrieved from: <https://www.aihw.gov.au/getmedia/c3e00f60-c40d-4989-ad22-de1be3ab5380/data-governance-framework-2021.pdf.aspx>.
6. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.* The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>
7. GOFAIR. Retrieved from: <https://www.go-fair.org/fair-principles/>.
8. Junior EPP, Normando P, Flores-Ortiz R, *et al.* Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South. *J Am Med Inform Assoc*. 2023 Mar 16;30(4):643–655. <https://doi.org/10.1093/jamia/ocac180>

9. Barreto M, Ichihara MY, Pescarini JM, *et al.* Cohort Profile: The 100 Million Brazilian Cohort. *International Journal of Epidemiology*, Volume 51, Issue 2, April 2022, Pages e27–e38. <https://doi.org/10.1093/ije/dyab213>
10. Pescarini JM, Williamson E, Ichihara MY, *et al.* Conditional Cash Transfer Program and Leprosy Incidence: Analysis of 12.9 Million Families From the 100 Million Brazilian Cohort. *American Journal of Epidemiology*, Volume 189, Issue 12, December 2020, Pages 1547–1558. <https://doi.org/10.1093/aje/kwaa127>
11. Ramos D, da Silva NB, Ichihara MY, *et al.* Conditional cash transfer program and child mortality: A cross-sectional analysis nested within the 100 Million Brazilian Cohort. *PLoS Med.* 2021 18(9): e1003509. <https://doi.org/10.1371/journal.pmed.1003509>
12. Alves FJO, Ramos D, Paixão ES, *et al.* Association of Conditional Cash Transfers With Maternal Mortality Using the 100 Million Brazilian Cohort. *JAMA Netw Open.* 2023 6(2):e230070. <https://doi.org/10.1001/jamanetworkopen.2023.0070>
13. World Health Organization. Social determinants of health. Retrieved from: [https://www.who.int/health-topics/social-determinants-of-health#tab=tab\\_3](https://www.who.int/health-topics/social-determinants-of-health#tab=tab_3)
14. World Health Organization. Health in all policies: Helsinki statement. Framework for country action (2014). Retrieved from: <https://www.who.int/publications/i/item/9789241506908>.
15. Grath-Lone LM, Jay MA, Blackburn R, *et al.* What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. *Int J Popul Data Sci.* 2022 Apr 27;7(1):1718. <https://doi.org/10.23889/ijpds.v6i1.1718>
16. Bashir H, Ferreira A, Ortigoza A, *et al.* Making the Invisible, Visible: Race, Racism, and Health Data, Lessons from Latin American Countries. The SALURBAL Project, the Ubuntu Center, and the Pan-DIASPORA Project. Drexel University Dornsife School of Public Health; January 2023. Retrieved from: <https://drexel.edu/lac/data-evidence/briefs/>.
17. Rebouças P, Goes E, Pescarini J, *et al.* Ethnoracial inequalities and child mortality in Brazil: a nationwide longitudinal study of 19 million newborn babies. *Lancet Glob Health.* 2022 Oct;10(10):e1453–e1462. [https://doi.org/10.1016/S2214-109X\(22\)00333-3](https://doi.org/10.1016/S2214-109X(22)00333-3)
18. Cardoso AM, Tavares IN, Werneck GL. Indigenous health in Brazil: from vulnerable to protagonists. *Lancet.* 2022 Dec 10;400(10368):2011–2014. [https://doi.org/10.1016/S0140-6736\(22\)02419-9](https://doi.org/10.1016/S0140-6736(22)02419-9)
19. Pita R, Sena S, Fiaccone R, *et al.* On the Accuracy and Scalability of Probabilistic Data Linkage Over the Brazilian 114 million Cohort. *IEEE Journal of Biomedical and Health Informatics.* 2018 Mar;22(2):346–353. <https://doi.org/10.1109/JBHI.2018.2796941>
20. Barbosa GCG, Ali MS, Araujo B, *et al.* CIDACS-RL: a novel indexing search and scoring-based record linkage system for huge datasets with high accuracy and scalability. *BMC Med Inform Decis Mak* 20, 289 (2020). <https://doi.org/10.1186/s12911-020-01285-w>
21. McGrail K, Jones K, Akbari A, *et al.* A Position Statement on Population Data Science: The science of data about people. *International Journal of Population Data Science.* 2018 3(1). <https://doi.org/10.23889/ijpds.v3i1.415>
22. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health.* 2011 32:91–108. <https://www.annualreviews.org/doi/10.1146/annurev-publhealth-031210-100700>
23. Cameron D, Pope S, Clemence M. Dialogue on Data: Exploring the public's views on using administrative data for research purposes. Ipsos MORI Social Research Institute, 2014. Retrieved from: <https://www.ipsos.com/sites/default/files/publication/1970-01/sri-dialogue-on-data-2014.pdf>
24. Waind, E. Trust, security and public interest: striking the balance A narrative review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research. *International Journal of Population Data Science.* 2020 5(3). <https://doi.org/10.23889/ijpds.v5i3.1368>
25. Stockdale J, Cassel J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. Wellcome Open Research, Jan 3(6), 2019. Retrieved from: <https://wellcomeopenresearch.org/articles/3-6/v2>.
26. Aitken M, de St. Jorre J, Pagliari C, *et al.* Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Medical Ethics.* 2016 17 (73). <https://doi.org/10.1186/s12910-016-0153-x>
27. Christen P, Schnell R. Thirty-three myths and misconceptions about population data: from data capture and processing to linkage. *International Journal of Population Data Science.* 2023 8:1:03. <https://doi.org/10.23889/ijpds.v8i1.2115>
28. SafePods: Secure Data Access. Retrieved from: <https://safepodnetwork.ac.uk/>
29. Boyd A, Flaig R, Oakley J, *et al.* The UK Longitudinal Linkage Collaboration: A trusted research environment for the longitudinal research community. *International Journal of Population Data Science.* 2023 8(2). <https://doi.org/10.23889/ijpds.v8i2.2299>
30. Mackey E, Elliot M, O'Hara K, Tudor C. The Anonymisation Decision-Making Framework, Manchester: UKAN, 2016. Retrieved from: <https://eprints.soton.ac.uk/399692/1/The-Anonymisation-Decision-making-Framework.pdf>

31. Shipsey, R, Plachta, J. Guidance: Linking with anonymised data – how not to make a hash of it. Office for National Statistics of the United Kingdom, updated 16 July 2021. Retrieved from: <https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/linking-with-anonymised-data-how-not-to-make-a-hash-of-it>
32. European Union. General Data Protection Regulation (GDPR). Regulation (EU) 2016/679. Retrieved from: <https://gdpr-info.eu/>
33. Britainthinks:Insight & Strategy. Improving lives through linked data: Views from groups with complex needs. Full report, March 2023. Retrieved from: [https://thinksinsight.com/wp-content/uploads/2023/03/BOLD\\_Public-Engagement\\_Full-report\\_Final-Publication-Version-23-03-23.pdf](https://thinksinsight.com/wp-content/uploads/2023/03/BOLD_Public-Engagement_Full-report_Final-Publication-Version-23-03-23.pdf)
34. Bennett CJ, Raab CD. Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulations & Governance*. 2018 1–18. <https://doi.org/10.1111/rego.12222>
35. Viljoen S. A Relational Theory of Data Governance . *Yale Law Journal*. 2020 Vol. 131 Retrieved from: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3727562](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562)
36. Milan S, Treré E. Big data from the South(s): Beyond data universalism. *Television & New Media*. 2019 20(4) 319–335. <https://doi.org/10.1177/1527476419837739>
37. Floridi L, Taddeo M. ‘What is data ethics?’ *Philosophical transactions of the Royal Society a Mathematical, Physical and Engineering Sciences*. 2016 374 (2083): 20160360. <https://doi.org/10.1098/rsta.2016.0360>
38. Rowe RK, Carrol SR, Healy C, et al. The SEEDS of Indigenous population health data linkage. *International Journal of Population Data Science*. 2021 6:1:22. <https://doi.org/10.23889/ijpds.v6i1.1417>
39. Carroll SR, Rodriguez-Lonebear D, Martinez A. Indigenous Data Governance: Strategies from United States Native Nations. *Data Sci J*. 2019 18:31. <https://doi.org/10.5334/dsj-2019-031>
40. The First Nations Information Governance Centre. First nations data sovereignty in Canada. *Statistical Journal of the IAOS*. 2019 35 (1), p. 47-69 . <https://doi.org/10.3233/SJI-180478>
41. Global Indigenous Data Alliance. Retrieved from: <https://www.gida-global.org/>.
42. Ada Lovelace Institute. Participatory Data Stewardship: A framework for involving people in the use of data, 2021. Retrieved from: [https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/content/uploads/2021/11/ADA\\_Participatory-Data-Stewardship.pdf](https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/content/uploads/2021/11/ADA_Participatory-Data-Stewardship.pdf)
43. ADRUK Public Engagement Strategy for 2021-2026. Retrieved from: [https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADR\\_UK\\_Public\\_Engagement\\_Strategy\\_2021-2026.pdf](https://www.adruk.org/fileadmin/uploads/adruk/Documents/ADR_UK_Public_Engagement_Strategy_2021-2026.pdf)
44. SAIL Databank. Consumer Panel. Retrieved from: <https://saildatabank.com/governance/approvals-public-engagement/consumer-panel/>.
45. SAIL Databank. Public Involvement & Engagement. Retrieved from: <https://saildatabank.com/governance/approvals-public-engagement/public-involvement-engagement/>.
46. Micheli M, Ponti M, Suman AB. Emerging models of data governance in the age of datification. *Big Data & Society*. 2020 July-December: 1–15. <https://doi.org/10.1177/2053951720948087>
47. Almeida BA, Pimenta DM. Perceptions and experiences on data sharing and linkage for research and the evaluation of public health policy, 2021. Full research report in Portuguese and executive summaries in Portuguese, English and Spanish available at: <https://cidacs.bahia.fiocruz.br/2021/11/22/percepcoes-e-experiencias-sobre-compartilhamento-e-vinculacao-de-dados-para-pesquisa-e-avaliacao-de-politicas-publicas-na-area-da-saude/>.
48. CIDACS, Public Engagement of Science. Available at: <https://cidacs.bahia.fiocruz.br/engajamento-publico-da-ciencia/>.
49. Dos Anjos Fonseca A, Pimenta DM, de Almeida MRS, et al. Public Involvement & Engagement in health inequalities research on COVID-19 pandemic: a case study of CIDACS/FIOCRUZ BAHIA. *Int J Popul Data Sci*. 2023 Jun 6;5(4):2133. <https://doi.org/10.23889/ijpds.v5i3.2133>

